STIC Search Report
Biotech-Chem Library

STIC Database Tracking Number: 94617

TO: Cynthia Collins **Location:** 9A12 **Art Unit: 1638** **Sunday, May 25, 2003** **Case Serial Number: 701926**	**From: Mary Jane Ruhl** **Location: Biotech-Chem Library** **CM1-6A06** **Phone: 605-1155** **maryjane.ruhl@uspto.gov**

Search Notes

12/4/00
6/4/98
9/25/98
6/4/99

GenCore version 5.1.5

OM nucleic - nucleic search, using sw model

Run on: May 24, 2003, 06:34:19 ; Search time 1707 Seconds
(without alignments)
11546.519 Million cell updates/sec

Title: US-09-701-926B-1
Perfect score: 1217
Sequence: 1 tttgaaatttatgtatttat.....gtcaacacaacacaacaaca 1217

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
c	1	62.2	5.1	1101	17	CNS00EVL	AL069706 Drosophil
	2	62	5.1	1101	17	CNS0039G	AL063921 Drosophil
	3	56.6	4.7	974	17	CNS00ITT	AL075432 Drosophil
	4	56	4.6	592	17	CNS005ER	AL059913 Drosophil
	5	55	4.5	1029	17	CNS01ZGM	AL174271 Tetraodon
c	6	55	4.5	1091	17	CNS014AC	AL103902 Drosophil
c	7	55	4.5	1101	17	CNS0039G	AL063921 Drosophil
	8	54.4	4.5	996	17	CNS00K06	AL076957 Drosophil
	9	53	4.4	1101	17	CNS0100X	AL098379 Drosophil
c	10	52.6	4.3	1204	13	BM452144	BM452144 AGENCOURT
c	11	52.4	4.3	1135	17	CNS033GQ	AL226115 Tetraodon
	12	52	4.3	872	17	AZ548700	AZ548700 ENTGE67TF
	13	52	4.3	1101	17	CNS000B8	AL063632 Drosophil
c	14	50.8	4.2	1092	17	CNS020K7	AL175696 Tetraodon
c	15	50.4	4.1	604	9	AL546530	AL546530 AL546530
c	16	50.4	4.1	959	17	CNS00655	AL062806 Drosophil
	17	50.4	4.1	1101	17	CNS0039R	AL063932 Drosophil
c	18	50.4	4.1	1101	17	CNS0106X	AL098595 Drosophil
c	19	50	4.1	857	17	AZ195387	AZ195387 SP_1030_A
	20	50	4.1	942	17	CNS018GS	AL109318 Drosophil
	21	50	4.1	1000	17	CNS00COQ	AL059446 Drosophil
	22	49.6	4.1	725	17	BH180166	BH180166 016_I_02-
	23	49.6	4.1	1204	17	CNS016E2	AL106628 Drosophil
	24	49.4	4.1	964	17	CNS07E8R	AL441457 T7 end of
c	25	49.2	4.0	938	17	CNS011LO	AL100422 Drosophil
	26	49	4.0	919	17	AZ535763	AZ535763 ENTCI26TF
	27	49	4.0	1101	17	CNS016LI	AL106896 Drosophil
	28	48.8	4.0	600	17	CNS06G12	AL397116 T3 end of
c	29	48.8	4.0	928	17	CNS00DKY	AL071865 Drosophil
c	30	48.8	4.0	967	17	CNS00JUP	AL077063 Drosophil
c	31	48.6	4.0	972	17	CNS002AX	AL097443 Drosophil
c	32	48.4	4.0	639	17	CNS017QD	AL108367 Drosophil
	33	48.2	4.0	874	17	CNS013EH	AL102755 Drosophil
c	34	48.2	4.0	1055	14	BQ876453	BQ876453 AGENCOURT
	35	48	3.9	963	9	AL566565	AL566565 AL566565
	36	47.8	3.9	904	17	CNS009GS	AL053354 Drosophil
	37	47.8	3.9	912	17	B10497	B10497 T2A12-Sp6 T
	38	47.8	3.9	950	17	BH165843	BH165843 ENTSR28TR
	39	47.8	3.9	961	17	CNS00LE2	AL067999 Drosophil
	40	47.8	3.9	1167	17	CNS07360	AL427102 clone BA0
	41	47.8	3.9	1201	17	CNS01096	AL098676 Drosophil
	42	47.6	3.9	932	9	AL523242	AL523242 AL523242
	43	47.6	3.9	1101	17	CNS012TP	AL102007 Drosophil
c	44	47.4	3.9	804	17	AG077527	AG077527 Pan trogl
	45	47.4	3.9	827	17	CNS00H7F	AL073027 Drosophil

ALIGNMENTS

RESULT 1
CNS00EVL/c
LOCUS CNS00EVL 1101 bp DNA linear GSS 04-JUN-1999
DEFINITION Drosophila melanogaster genome survey sequence T7 end of BAC: BACR29B23 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL069706
VERSION AL069706.1 GI:4949849
KEYWORDS GSS.
SOURCE Drosophila melanogaster.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1101)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial

EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES Location/Qualifiers
source 1. .1101
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone="BACR29B23"
/clone_lib="RPCI-98"
/note="end : T7"

BASE COUNT 419 a 91 c 60 g 299 t 232 others
ORIGIN

Query Match 5.1%; Score 62.2; DB 17; Length 1101;
Best Local Similarity 33.7%; Pred. No. 0.096;
Matches 204; Conservative 92; Mismatches 305; Indels 4; Gaps 1;

```
Qy    163 TATCACTGATGGTGATTCAGTATTGTGATTATGTCCTTCGTTGATTATTCTGTTTCATAC 222
Db   1100 TWWTATWTWTTTWTTTTTWWTATWTTAWATATATAAWTAATTWWTWTATATWTAWWWTAT 1041

Qy    223 AAGTCGTGTAATTTGCTGTTTGTGACAGTACGATAGATCGACTCAACCTTCTGAGGTATT 282
Db   1040 WATATATTAWTTTWTAAAATATAWWTTAATAAATATATWTAWTAWWATATATAWATAWAA 981

Qy    283 AGTTGAAGTTCATGTAAATTAGCTTTGTTTATCATAGTAGCATTTGATTATTGATGCTCT 342
Db    980 WTATAWAWTWWAATWAATWWATAWAWTTTAATAAWAATATWAWWWTWTTATTTTTTTTTT 921

Qy    343 GTAGCTAATGATAAGCCATTGGAGGGAAGCAAGCTTTCTAAATGAATCTACGAATGGATG 402
Db    920 TTTWTTTWTATWTWTTTATATWWTWWAAAAWAWWAAWTWTATTTTTTWTTATWATTWATA 861

Qy    403 ATAAAGTTCATGAATATTTTTGTTACTTCTGCAGTCAGATCATGAGTTATTGAGTCTATT 462
Db    860 TTAWATTTTTTWTWTATTWWTATATWATWWTATWT----WTWTTTWATWTATWTATTTAT 805

Qy    463 GTTTTTTTAAGCCTGTTTCAGATGATCCATCATCAGTAACAACATACACGGTGTAGTCCC 522
Db    804 WTWTWTTWTATTWTATWTATATAWWWTWWATWWWTAWAAWTATATATWWATATTATWWTWAT 745

Qy    523 AAATCCATCATATGCACCTTCTTTTCTTCAATTTGGTCTTGTTTTTTTTTTTTCATGATG 582
Db    744 TWTTWTTTTATTWTTTTAAWTTATWTATWATTWTTWATWTTTTTTTTTTTTTTTWTATTT 685

Qy    583 TCATTGAATTATTCAAGAAGTCACTTCGAGCATAATGATTTTTCAAAATCCACCTTTGTT 642
Db    684 WAAATTATTTAATTAWWWAATWAWTTATTTAATAAWTAATTAAATTATTAAAAATTTTAT 625

Qy    643 CAAGCACTACCACGTCTTTTCATCTAGCCCACAACCGTGGTGGAGGATCTAGAATTTTCA 702
Db    624 WAAAAAATATTTTTTTATATTAATWAWATTTATAWATTAATATAWWTTWTAAATTWTTAA 565

Qy    703 TGAAAGGATTCAAAATTTACAAACATATATATACACTATACACTATGAATCCACTAATAC 762
Db    564 TWTATTWWTTTAAATTATWTTTTTTTTTAWTAWAWWTWAAAWAATWAWWWWWAAAAATWW 505

Qy    763 TAGAT 767
Db    504 WWTWT 500
```

RESULT 2
CNS0039G
LOCUS CNS0039G 1101 bp DNA linear GSS 03-JUN-1999
DEFINITION Drosophila melanogaster genome survey sequence TET3 end of BAC # BACR08K10 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL063921
VERSION AL063921.1 GI:4941778
KEYWORDS GSS.
SOURCE Drosophila melanogaster.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1101)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES Location/Qualifiers
source 1. .1101
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone="BACR08K10"
/clone_lib="RPCI-98"
/note="end : TET3"
BASE COUNT 201 a 64 c 131 g 202 t 503 others
ORIGIN

Query Match 5.1%; Score 62; DB 17; Length 1101;
Best Local Similarity 15.6%; Pred. No. 0.1;
Matches 81; Conservative 235; Mismatches 202; Indels 0; Gaps 0;

```
Qy    532 ATATGCACCTTCTTTTCTTCAATTTGGTCTTGTTTTTTTTTTTTTCATGATGTCATTGAAT 591
Db    542 ATWTWWTWWTTYTTWAAWATAAAMCMAAWYYHTYTTYTYHYYTYWTYTMTWHYHTMYTHA 601

Qy    592 TATTCAAGAAGTCACTTCGAGCATAATGATTTTTCAAAATCCACCTTTGTTCAAGCACTA 651
Db    602 WAHTTWYHWYHTYAMWHWMTWHWTMWAWWHWTTYTAAYYYYYTCMYYYHYMHWHHAHAHA 661

Qy    652 CCACGTCTTTTCATCTAGCCCACAACCGTGGTGGAGGATCTAGAATTTTCATGAAAGGAT 711
Db    662 AAWWTTHTWWTHAYHWATYHYYYYMYCAMMCMCTHTCHHCYYYYHHYTAHHTHTHHWYAH 721

Qy    712 TCAAAATTTACAAACATATATATACACTATACACTATGAATCCACTAATACTAGATGGTG 771
Db    722 YYMWYYMWAYYWMYCTACTYHYHHHHHYHWAYHTTWYAWAHAMWMWHHAHYAAAAAWAAW 781

Qy    772 CACCTGTGCCCCCACTCATGTGAAAGCCTATTCTCAATTTTTTATTTTCCACAACTTAAA 831
Db    782 ATTHHYHHTTHYMHHTYMYHYYMYTCCYMCTYHCWHYYHTAYTCWTWTHHWMWTWTHWYH 841

Qy    832 TACAGACCGCACAACTCCCGTGTCTTGTGTGCTCGTCGCTCAGCATGCAAGTCGAGAAAA 891
Db    842 HTWWHHTTTHWAWWHTHTWCWWWWHATTWTWATHCWACMTMHWHHWMHMHHHHMACHAHH 901

Qy    892 GAAAGACCAAAACAATGAAAACTTTACGAAAAATCAAAAAGTTGAAGGACTTTAACGTCG 951
Db    902 HTHMCMCHHHMHCTCHHHHTMYHMTCHWMWHMHWHWHHWWMATWMTTMTTMMMMCCMMHH 961

Qy    952 AGATCTCTCGTAGAAAACCTCTTTTGTAAGGTTGCATACAATACTTTTTTTTCAGACTTT 1011
Db    962 HCHMYHMMHMYMYCCHYYCTCHTHATTHYHYMCTCYHYCTWHTYWTAYWWAWTAHAMTTA 1021
```

```
Qy   1012 ACTTATGGTATTATACTGAATATGTTATTGCTGTTATA 1049
          :::   :|::||: : ::::   ::   |: :|:::
Db   1022 TWWWWMHWWAHWATWWWWWWWWATAWWACTCHHWTWYH 1059


RESULT 3
CNS00ITT
LOCUS       CNS00ITT                  974 bp    DNA     linear   GSS 03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence TET3 end of BAC:
            BACR37D06 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL075432
VERSION     AL075432.1  GI:4954990
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 974)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .974
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone="BACR37D06"
                     /clone_lib="RPCI-98"
                     /note="end : TET3"
BASE COUNT       77 a     59 c     13 g    605 t    220 others
ORIGIN

  Query Match             4.7%;  Score 56.6;  DB 17;  Length 974;
  Best Local Similarity   32.2%;  Pred. No. 1;
  Matches  264;  Conservative  100;  Mismatches  455;  Indels    0;  Gaps    0;

Qy       1 TTTGAAATTTATGTATTTATCTATAGCATTAGAAACTATAAGAGTTGTTAGCTTCACTTG 60
             ||     ||| | | | | |         ||       | |    || ||   |    ||
Db      70 TTNTTTNTTTTTNTTTNTTTTNTNTTTTTTTTTTNTTTTGTGTTTTTTTTNTNTNTTTT 129

Qy      61 GCTTACTGTTGTGCTCAAAGCAACTTCATCATCATACAGTATGGTTTTGATATGCTCTTC 120
             ||  | || |  |       : ||  |  |  |    | |  ||||  | |  | ||
Db     130 TTTTTTTTTTTTTTTTTTTTYRTRTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 189

Qy     121 CATTATCACTGAGCCTTATGATTATGTTTTACGAGCTTATAATATCACTGATGGTGATTC 180
             || |   |     || |  || | ||||      || |  | |   |  | |  ||
Db     190 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 249

Qy     181 AGTATTGTGATTATGTCCTTCGTTGATTATTCTGTTTCATACAAGTCGTGTAATTTGCTG 240
             | || |  |  | |  ||  ||  || || | |||  |     |  | |  |||  |
Db     250 TTTTTTTTTTNTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 309

Qy     241 TTTGTGACAGTACGATAGATCGACTCAACCTTCTGAGGTATTAGTTGAAGTTCATGTAAA 300
             ||| |     ||:   | : :: | :   | |   ::|  |   ::::   : : :::
Db     310 TTTTTTTTTNTAYACAATWGTKVTTTTMTTNTTTTTTRHAARAASAVMMSMGRVVVSVVM 369

Qy     301 TTAGCTTTGTTTATCATAGTAGCATTTGATTATTGATGCTCTGTAGCTAATGATAAGCCA 360
           |:: :    ||: |  | : : ::||| : ::|: |    :|:|:  |  | : : :  : |
Db     370 TWVTYMVVTTTHTTTTTGSSVNVVTTTTWCWRTWTAAMNVSTVTMTATTTTTMAWTMMKA 429

Qy     361 TTGGAGGGAAGCAAGCTTTCTAAATGAATCTACGAATGGATGATAAAGTTCATGAATATT 420
           ||       | : ||: |   |  :  |  |||   |  | |  |  :  |:||
Db     430 TTWTTTTTTTMTATSTTTWTTTTTTHTWMATTWMAATATTTTSTTATATANMNAAMTACN 489

Qy     421 TTTGTTACTTCTGCAGTCAGATCATGAGTTATTGAGTCTATTGTTTTTTTAAGCCTGTTT 480
           ||| ||  || |   :||   ||   ::||: |    |   | |||||:::  : :  |
Db     490 TTTTTTTTTTTTTTTBTCTTTTCTVTMRTTMGTANMRCRTATTTTTTTTHHMMTMTWTMTC 549

Qy     481 CAGATGATCCATCATCAGTAACAACATACACGGTGTAGTCCCAAATCCATCATATGCACC 540
           |     |::  : |   |        |        |: : :   : | : |||:| ::
Db     550 CTMTMTAYMMCAMTTTTTTTTTTTTTTTTTTTTTMYTHTYTSASTMMCAMRCATWTAMMWA 609

Qy     541 TTCTTTTCTTCAATTTGGTCTTGTTTTTTTTTTTTTCATGATGTCATTGAATTATTCAAGA 600
           :: :|  :| : ||:|  | || ||||||||||||  |  | | :|   | | |
Db     610 KWTHTMMMTAMCATWTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTMTCNCACTTTVTTTTT 669

Qy     601 AGTCACTTCGAGCATAATGATTTTTCAAAATCCACCTTTGTTCAAGCACTACCACGTCTT 660
             |    |  : : |     |:|:|:    |     ||| ||: |    |:    : ||
Db     670 TTTTSNATTTWTMNTTTCTTTHTHTYTTTTTTTTTTTTTTTTTTMGAATTTTWTTCGAHTTT 729

Qy     661 TTCATCTAGCCCACAACCGTGGTGGAGGATCTAGAATTTTCATGAAAGGATTCAAAATTT 720
           : | : |     |||: :    |  :  |: :  ::||:  :|   |   |: :::|
Db     730 WMCTWTTTTTTTACAMTMNNMTTTTWATABTWTTWMTTWMTWTWYTACAYTHTVWMACRC 789

Qy     721 ACAAACATATATATACACTATACACTATGAATCCACTAATACTAGATGGTGCACCTGTGC 780
           |  |  : |  |||  :    ||     ||    :| |:|  :|      |
Db     790 AACACACWTTTMKTACTWHTAYTACACNMCATAATAYACTRCACAMTWTACTTTCCWRTT 849

Qy     781 CCCCACTCATGTGAAAGCCTATTCTCAATTTTTTATTTT 819
           | :|| |  |   |    |.::||    |||||| ||||
Db     850 CTMCANTTTTTCTACTYWCMMYTCCTTTTTTTTTTTTTT 888


RESULT 4
CNS005ER
LOCUS       CNS005ER                  592 bp    DNA     linear   GSS 03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence TET3 end of BAC #
            BACR11E05 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL059913
VERSION     AL059913.1  GI:4943035
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 592)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
```

```
FEATURES             Location/Qualifiers
     source          1. .592
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone="BACR11E05"
                     /clone_lib="RPCI-98"
                     /note="end : TET3"
BASE COUNT      180 a     15 c     29 g    209 t    159 others
ORIGIN

  Query Match             4.6%;  Score 56;  DB 17;  Length 592;
  Best Local Similarity   41.7%;  Pred. No. 1.6;
  Matches  128;  Conservative   15;  Mismatches  164;  Indels    0;  Gaps    0;

Qy     880 AAGTCGAGAAAAGAAAGACCAAAACAATGAAAACTTTACGAAAAATCAAAAAGTTGAAGG 939
           ||    | |||| ||| |  ||| ||  ||||    | |||||  |||||    ||
Db      13 AANAAAAAAAAAAAAAAAAAAAAAANAAANNAAAAAAAAAAAAAAAAAAAAAAAAAAAA 72

Qy     940 ACTTTAACGTCGAGATCTCTCGTAGAAAACCTCTTTTGTAAGGTTGCATACAATACTTTT 999
           |   |||     | |  | |   | |||| | ||||      ||   |  |||  ||||
Db      73 AAAATAAAAATAATAAATTTATAATAAAAAATTTTTTTCTTTTTTTTTTTTTAATTTTTT 132

Qy    1000 TTTTCAGACTTTACTTATGGTATTATACTGAATATGTTATTGCTGTTATAGTAGTTGAGT 1059
           ||||     |||: :::|  :::| |  |   | | || ||  | ||  | |  ||
Db     133 TTTTTTTTTTTTWTWWWTTWWWWTTTTTTTTTTTTTTTTTTTTTTTTAATTTATTTTAN 192

Qy    1060 GACGTTTGAGGGAATTTCTAGTCCGTTAATCTTGTACTCAGTGTGTCTACTTTTCAAAAA 1119
             |    |     | |   |      ||  | || |         | |  |||| |  ||
Db     193 TANNNTNNNAATANTAATTTNATTTTTTTTTTTTTTWATTTATATTTTTTTTTTAANNAA 252

Qy    1120 AGTCAGTTTTTCAGTCTCTAAAACACATTTAAATAAGAGTTTCTTTGCCCATCTTTTGTT 1179
           |   : |:||  |      |||| |  |  | :| |   ||: ||     |:  |   :
Db     253 AAAAWNTWTTAAATNNNNNAAAAAADTTANATWTTATTTTTKNTTANTTNAWNATCNNNY 312

Qy    1180 CCTCATC 1186
           ::|  ||
Db     313 SMTTCTC 319


RESULT 5
CNS01ZGM
LOCUS       CNS01ZGM                1029 bp    DNA     linear   GSS 12-MAY-2000
DEFINITION  Tetraodon nigroviridis genome survey sequence PUC-Ori end of clone
            220I12 of library G from Tetraodon nigroviridis, genomic survey
            sequence.
ACCESSION   AL174271
VERSION     AL174271.1  GI:7812328
KEYWORDS    GSS; genome survey sequence.
SOURCE      Tetraodon nigroviridis.
  ORGANISM  Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetraodontidae; Tetraodon.
REFERENCE   1  (bases 1 to 1029)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
            Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
            Saurin,W. and Weissenbach,J.
  TITLE     Human gene number estimate provided by genome wide analysis using
            Tetraodon nigroviridis DNA sequence
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1029)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
            Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
            Weissenbach,J.
  TITLE     Charaterization and repeat analysis of the compact genome of the
            freshwater pufferfish Tetraodon nigroviridis
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 1029)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2000)
COMMENT     This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.
FEATURES             Location/Qualifiers
     source          1. .1029
                     /organism="Tetraodon nigroviridis"
                     /db_xref="taxon:99883"
                     /clone="220I12"
                     /clone_lib="G"
                     /note="Genoscope sequence ID : C0AG220BE06SP1-end :
                     PUC-Ori"
BASE COUNT      297 a    124 c     66 g    457 t     85 others
ORIGIN

  Query Match             4.5%;  Score 55;  DB 17;  Length 1029;
  Best Local Similarity   41.1%;  Pred. No. 1.9;
  Matches  223;  Conservative   28;  Mismatches  290;  Indels    2;  Gaps    1;

Qy      83 ACTTCATCATCATACAGTATGGTTTTGATATGCTCTTCCATTATCACTGAGCCTTATGAT 142
           | ||| | ||       | |  |||| || |  |  |   || |  |      || | ||
Db     412 ATTTCTTTATATATTTATTTTTTTTTAATTTTTTTATTTTTTTTTTTWTTTTTTTTTTAT 471

Qy     143 TATGTTTTACGAGCTTATAATATCACTGATGGTGATTCAGTATTGTGATTATGTCCTTCG 202
           | | ||||      || |  | |   |  |  |  ||   | || |  || | |  ||
Db     472 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTNTTTTTTTTTTTTTTTTTT 531

Qy     203 TTGATTATTCTGTTTCATACAAGTCGTGTAATTTGCTGTTTGTGACAGTACGATAGATCG 262
           ||  || || | |||  |     |   |  |||     |       |      || ||
Db     532 TTTTTTTTTTTTTTTTTTTTTTTTTTATTTTTTTATATATATATTTATATTTTTATATAT 591

Qy     263 ACTCAACCTTCTGAGGTATTAGTTGAAGTTCATGTAAATTAGCTTTGTTTATCATAGTAG 322
           |   ||   |   |  || | |  |  || || | : ||   ||| ||||| | | |:
Db     592 AAATAAAAATAAAAAAAATAATTATATATTAATTTTWTTTTTTTTTTTTTTATTAAATTWT 651

Qy     323 CATTTGATTATTGATGCTCTGTAGCTAATGATAAGCCATTGGAGGGAAGCAAGCTTTCTA 382
             :|  | || |  |      |  | |  |       | |      ||  ||      :|
Db     652 TTWTATAATAATTTTAAAAATTTTTTTAAAAATTTTAAGTATTAAAAAAAAAAWAAAAWA 711

Qy     383 AATGAATCTACGAATGGATGATAAAGTTCATGAATATTTTTGTTACTTCTGCAGTCAGAT 442
           ||  | | |  |||   | :| :| |: ::  ||||| :| ||: ||    |   :  |
Db     712 AAATATTTTTATAATTTTTTTWTTWAWTWTWWATATATTAWTTTTWTTTTAAAATATWTTT 771

Qy     443 CATGAGTTATTGAGTCTATTGTTTTTTTAAGCCTGTTTCAGATGATCCATCATCAGTAAC 502
           |  |||  |   :| | ||| ||| :   |  |  |  |  |   : || |  ||
Db     772 TTATAAATATATAAAWWAATATTTATTTTWATATWATAWATTTTTTTTWTWAATAATAAAA 831

Qy     503 AACATACACGGTGTAGTCCCAAATCCATCATATGCACCTTCTTTTCTTCAATTTGGTCTT 562
           |  ||| :   | : :|    |      |  || |: |  |  :|:| |: ||| |   ||
Db     832 ATAATAAWATTTTWTKTTTWATTATTAATATNTRTATATATWTWTATWTAATAT--ATTT 889

Qy     563 GTTTTTTTTTTTTCATGATGTCATTGAATTATTCAAGAAGTCACTTCGAGCATAATGATT 622
           | |||||||||| |: :    |:  :|||| | || || |   ||  |  |:|||  :|
Db     890 TTATTTTTTTTTTAAWAWATAWAWATWATTAATAAAAAAATTTWTTAAAAAAWAATATWT 949

Qy     623 TTT 625
           |||
Db     950 TTT 952


RESULT 6
CNS014AC/c
LOCUS       CNS014AC                1091 bp    DNA     linear   GSS 26-JUL-1999
DEFINITION  Drosophila melanogaster genome survey sequence SP6 end of BAC
            BACN11F13 of DrosBAC library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL103902
VERSION     AL103902.1  GI:5615513
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
```

```
REFERENCE   1  (bases 1 to 1091)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the European Drosophila Genome Project (EDGP) -
            http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
            library (Dros BAC) was made by Alain Billaud at CEPH (Centre
            d'Etude du Polymorphisme Humain) with funding provided by a MRC
            project grant. The DNA was prepared from embryos by Alain Bucheton
            and Genevieve Payan. It has been constructed in the vector
            pBeloBAC11.
FEATURES             Location/Qualifiers
     source          1. .1091
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone="BACN11F13"
                     /clone_lib="DrosBAC"
                     /plasmid="pBeloBAC11"
                     /note="end : SP6"
BASE COUNT      569 a     70 c     83 g    226 t    143 others
ORIGIN

  Query Match             4.5%;  Score 55;  DB 17;  Length 1091;
  Best Local Similarity   37.7%;  Pred. No. 1.8;
  Matches  208;  Conservative   36;  Mismatches  307;  Indels    0;  Gaps    0;

Qy      78 AAGCAACTTCATCATCATACAGTATGGTTTTGATATGCTCTTCCATTATCACTGAGCCTT 137
           | ||||||   |  |        | |  | |  : || || |     |  | | |     ||
Db     665 ATGCAACTATTTTNTNTGTTTTTNTTTTNTATWNATNCTNTNTTTTNTTWATTATWTNTT 606

Qy     138 ATGATTATGTTTTACGAGCTTATAATATCACTGATGGTGATTCAGTATTGTGATTATGTC 197
           |:   | | | ||:      ||||  | | | | ||  | :||   | ||    ||  | |
Db     605 AWTTCTTTTTCTTWTNTNTTTATTTTTTTAATAATTATAWTTTTTTTTTTTNTATATTTTA 546

Qy     198 CTTCGTTGATTATTCTGTTTCATACAAGTCGTGTAATTTGCTGTTTGTGACAGTACGATA 257
           ||  || :|| || | ||:  |      |  | || |||  | ||| | |    |     |
Db     545 TTTTNTTTWTTTTTTTTATTWTTTTTTTTTWTTTTATTTTTWTTTTTTTTATTTTTATTTT 486

Qy     258 GATCGACTCAACCTTCTGAGGTATTAGTTGAAGTTCATGTAAATTAGCTTTGTTTATCAT 317
           |    |     :| :    | ||  ||    :| |: |   ||| ||||  |:||  :
Db     485 TTTTTTNTTTTTTWTWWTTTTTTTTTTTTTTTTTWTTAWTTTCTTTACCTTTWATWATTTW 426

Qy     318 AGTAGCATTTGATTATTGATGCTCTGTAGCTAATGATAAGCCATTGGAGGGAAGCAAGCT 377
           |    ||:  ||:||  |    | |   |      :|| :  :  |    |
Db     425 TTTTTTTTTWATTTWTTTTTWTATTTTTTTTTTATTTTTTTTWTTTTWTWTWTTCTTTTT 366

Qy     378 TTCTAAATGAATCTACGAATGGATGATAAAGTTCATGAATATTTTTGTTACTTCTGCAGT 437
           || |   |  ||| |   |   | |    ||    | | ||||| ||  || |    |
Db     365 TTTTTTTTCCCTCTTCCYTTTTTTTTTTCTWTTTTCCATTTTTTTTTTTTTTTTTTTTTT 306

Qy     438 CAGATCATGAGTTATTGAGTCTATTGTTTTTTTAAGCCTGTTTCAGATGATCCATCATCA 497
           |  |   |: ||   | | :| |||| :| :   | ||:    |  |      | |
Db     305 TTTTTTTTTTTCTWTTTTCTTTTTWTTTTTTAWTTWTTWTTTTYWTTTTTTTTTWTCTACGA 246

Qy     498 GTAACAACATACACGGTGTAGTCCCAAATCCATCATATGCACCTTCTTTTCTTCAATTTG 557
           | |      : ||      |:      |   |           |: |  :  || | :||
Db     245 GCACATCANWTCAGATATCAKATAACAGAAAAGGTCCAAGGGCWGCAAYAATTTATWTTT 186

Qy     558 GTCTTGTTTTTTTTTTTTCATGATGTCATTGAATTATTCAAGAAGTCACTTCGAGCATAA 617
           ||  :|||||||:|| : :|   :|| | || |:       |||:||    |:||
Db     185 AAWTTTAWTTTTTTTTWTCTWTWTTATWTTWATTTCTWTTTTTTWTCAYTTTTTTCWTAT 126

Qy     618 TGATTTTTCAA 628
           :  ||| ||||
Db     125 WTYTTTATCAA 115


RESULT 7
CNS0039G/c
LOCUS       CNS0039G                1101 bp    DNA     linear   GSS 03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence TET3 end of BAC #
            BACR08K10 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL063921
VERSION     AL063921.1  GI:4941778
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .1101
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone="BACR08K10"
                     /clone_lib="RPCI-98"
                     /note="end : TET3"
BASE COUNT      201 a     64 c    131 g    202 t    503 others
ORIGIN

  Query Match             4.5%;  Score 55;  DB 17;  Length 1101;
  Best Local Similarity   16.4%;  Pred. No. 1.8;
  Matches  103;  Conservative  272;  Mismatches  249;  Indels    3;  Gaps    2;

Qy       1 TTTGAAATTTATGTATTTATCTATAGCATTAGAAACTATAAGAGTTGTTAGCTTCACTTG 60
           | :::::::|::: :::: |:  ::|| |: : :::   |:: | :: ::   | : :::
Db    1087 TRKDDWDWTKWWTWWKDRADDRRWAGDADRWAWDDGAGTWWTATWWWWWWWWATWDTWWD 1028

Qy      61 GCTTACTGTTGTGCTCAAAGCAACTTCATCATCATACAGTATGGTTTTGATATGCTCTTC 120
           : ::: |  : :  : :::  :: ::     :   :: :: ||: :   ::: :   ::
Db    1027 KWWWWATAAKTDTAWTWWRTAWRADWAGRDRGAGKRDRDAATDADGAGRRDGGRKRKDKK 968

Qy     121 CATTATCACTGAGCCTTATGATTATGTTTTACGAGCTTATAATATCACTGATGGTGATTC 180
           ::: : : :| :  : |: |::||::::::  :: :::: |::  | :::  |:::::
Db     967 DRKDGDDDKKGGKKKAAKAAKWATKWWDDWDWDKDWKWDGAKDRKADDDDGAGDKDDDG 908

Qy     181 AGT-ATTGTGATTATGTCCTTCGTTGATTATTCTGTTTCATACAAGTCGTGTAATTTGCT 239
           |: |:::|: |:::::  :: ::::|:  :     : : ||: ::    : :::|::
Db     907 KGKDADDDTDGTKDDDDKDKWDDWDKAKGTWGDATWAWAATDWWWWGWADADWWTWDAAA 848

Qy     240 GTTTGTGACAGTACGATAGATCGACTCAACCTTCTGAGGTATTAGTTGAAGTTCATGTAA 299
           :::: ::: : :| :::: :  || | ::  :      ::      : :: : :: ::  ::
Db     847 DDWWADDRWDAWAWKWDDAWAWGARTADRRDWGDRAGKRGGARKRRDRKRADDKRDAADD 788

Qy     300 ATTAGCTTTGTTTATCATAGTAGCATTTGATTATTGATGCTCTGTAGCTAATGATAAGCC 359
           :::|  :|| ||| | :|:::    :|: :: |:  ::: : :: ::   | : :::
Db     787 RDDAATWTTWTTTTTRDTDDWKWKTDTWTRWAADRTWDRDDDDDRDRAGTAGRKWRRTWK 728

Qy     360 ATTGGAGGGAAGCAAGCTTTCTAAATGAATCTACGAATGGATGATAAAGTTCATGAATAT 419
           : ::::: :::: :|: |   :::: |::  :| :  : |:::: ::: |: :|:|::|:
Db     727 RRWKRRDTRWDDADADDTARDDRRRRGDDGADAGKGKKTGRKRRRRDRATWDRTDAWWAD 668
```

```
Qy      420 TTTTGTTACTTCTGCAGTCAGATCATGAGTT--ATTGAGTCTATTGTTTTTTTTAAGCCTG 477
Db      667 AAWWTTTDTDTDDWDKRDRRRKGARRRRRTTARAAWDWWTWKAWDWAKWDWKTRADRWDR 608

Qy      478 TTTCAGATGATCCATCATCAGTAACAACATACACGGTGTAGTCCCAAATCCATCATATGC 537
Db      607 WAADTWTDARKADRDWAKARAWRARRDRARAARADRRWTTKGKTTTATWTTWAARAAWWA 548

Qy      538 ACCTTCTTTTCTTCAATTTGGTCTTGTTTTTTTTTTTTCATGATGTCATTGAATTATTCA 597
Db      547 WWAWATTTATWTTTWTTTWTTTTTTTTTTTTWTTTAAWWWAAWWTATWAAWTAAAWAAAAA 488

Qy      598 AGAAGTCACTTCGAGCATAATGATTTT 624
Db      487 AAWAATTTTTTTTTTTTWAWTAAAWWT 461
```

RESULT 8
CNS00K06
LOCUS CNS00K06 996 bp DNA linear GSS 03-JUN-1999
DEFINITION Drosophila melanogaster genome survey sequence TET3 end of BAC: BACR39E23 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL076957
VERSION AL076957.1 GI:4956535
KEYWORDS GSS.
SOURCE Drosophila melanogaster.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 996)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES Location/Qualifiers
source 1. .996
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone="BACR39E23"
/clone_lib="RPCI-98"
/note="end : TET3"
BASE COUNT 226 a 109 c 110 g 418 t 133 others
ORIGIN

Query Match 4.5%; Score 54.4; DB 17; Length 996;
Best Local Similarity 36.6%; Pred. No. 2.4;
Matches 194; Conservative 51; Mismatches 285; Indels 0; Gaps 0;

```
Qy       45 TTGTTAGCTTCACTTGGCTTACTGTTGTGCTCAAAGCAACTTCATCATCATACAGTATGG 104
Db      364 KTTTTTTTYTTTCTTGTTTYTGGATTTTTTTGTTTGGTTTCWTWTTTTCGGTCAATTTKG 423

Qy      105 TTTTGATATGCTCTTCCATTATCACTGAGCCTTATGATTATGTTTTACGAGCTTATAATA 164
```

```
Db      424 GTTTTTTYTTTCGWTWTAYTGTTTCTGGAATTTTTTCTKTCACTTTTGTTTCCWTCTTGT 483

Qy      165 TCACTGATGGTGATTCAGTATTGTGATTATGTCCTTCGTTGATTATTCTGTTTCATACAA 224
Db      484 TCTTTTTGGGGATTTTTTGAGTTTGGTNCTCCGGTTCTCGTTGTTTTCTTTTTTTTWCAG 543

Qy      225 GTCGTGTAATTTGCTGTTTGTGACAGTACGATAGATCGACTCAACCTTCTGAGGTATTAG 284
Db      544 AATAWWWTWTTTKWTTTTWTTTWTWACAATAACCAACAATTTATTATTTTTTYYWWATTK 603

Qy      285 TTGAAGTTCATGTAAATTAGCTTTGTTTATCATAGTAGCATTTGATTATTGATGCTCTGT 344
Db      604 TTTAAGGAGATAAAATCACTTTCCGATCACTTTWWWWTYATTTTTTTTTWTTWWKAATTT 663

Qy      345 AGCTAATGATAAGCCATTGGAGGGAAGCAAGCTTTCTAAATGAATCTACGAATGGATGAT 404
Db      664 MCGWTAMGAAAAMAWMGNATAMGATTATTTWYATTTTTYTTACAAAGCCCTATTKATACT 723

Qy      405 AAAGTTCATGAATATTTTTGTTACTTCTGCAGTCAGATCATGAGTTATTGAGTCTATTGT 464
Db      724 ATCACACWAGCCAAAATTTATTTTTTYWWCACACAAAWTTTTTTTTTTTKTTGTCATTAWW 783

Qy      465 TTTTTTAAGCCTGTTTCAGATGATCCATCATCAGTAACAACATACACGGTGTAGTCCCAA 524
Db      784 GWTTTTTTTTWTYTYGAACAACAACCTGTAATWATWGCAGACCAAATGGAMATMHMMAYA 843

Qy      525 ATCCATCATATGCACCTTCTTTTCTTCAATTTGGTCTTGTTTTTTTTTTT 574
Db      844 TTMAAAATTGWATTGCATAAAGAAMWMAWTTTTTTTTTTTTTTTTTTTTY 893
```

RESULT 9
CNS0100X
LOCUS CNS0100X 1101 bp DNA linear GSS 26-JUL-1999
DEFINITION Drosophila melanogaster genome survey sequence SP6 end of BAC BACN03G04 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL098379
VERSION AL098379.1 GI:5609990
KEYWORDS GSS.
SOURCE Drosophila melanogaster.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1101)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the European Drosophila Genome Project (EDGP) - http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC library (Dros BAC) was made by Alain Billaud at CEPH (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pBeloBAC11.
FEATURES Location/Qualifiers
source 1. .1101
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone="BACN03G04"
/clone_lib="DrosBAC"
/plasmid="pBeloBAC11"
/note="end : SP6"
BASE COUNT 195 a 108 c 131 g 161 t 506 others
ORIGIN

Query Match 4.4%; Score 53; DB 17; Length 1101;
Best Local Similarity 19.8%; Pred. No. 4.1;
Matches 102; Conservative 185; Mismatches 229; Indels 0; Gaps 0;

```
Qy     115 CTCTTCCATTATCACTGAGCCTTATGATTATGTTTTACGAGCTTATAATATCACTGATGG 174
Db     312 CNNNNNCCTTTTNTNTTTTNTNNTBDADGWWATARWAWDCRYATATAAAADAACTTWTKT 371

Qy     175 TGATTCAGTATTGTGATTATGTCCTTCGTTGATTATTCTGTTTCATACAAGTCGTGTAAT 234
Db     372 TTTTGTDTDKTTTKTDTKTTTTRTTKGGGGTGKRTGNGTGTGAGKGDGKGDGGKGAKWTD 431

Qy     235 TTGCTGTTTGTGACAGTACGATAGATCGACTCAACCTTCTGAGGTATTAGTTGAAGTTCA 294
Db     432 TRTDDTWKTDTDTWGKADTDWWTDTWKTTWTWTWTTTTGWKKWDTKTKTKWKDDDGTTWT 491

Qy     295 TGTAAATTAGCTTTGTTTATCATAGTAGCATTTGATTATTGATGCTCTGTAGCTAATGAT 354
Db     492 DTWKDWKTDKDTDKTWTDVWADNAGWWTDKDKTKAKKKRTKTDTKKTKTRTGTKKKTDRG 551

Qy     355 AAGCCATTGGAGGGAAGCAAGCTTTCTAAATGAATCTACGAATGGATGATAAAGTTCATG 414
Db     552 MRKTDTWGWGDWRADRTDGARADRAANAKDKDADGKADTDATKTDGWGTTTKADTTDDTD 611

Qy     415 AATATTTTTGTTACTTCTGCAGTCAGATCATGAGTTATTGAGTCTATTGTTTTTTTAAGC 474
Db     612 ADWWDWKDWDWKDADTKAGRKRGWGWGKKGGWKKTTKTKKKKTDKTKTGTDTKWGWKDKW 671

Qy     475 CTGTTTCAGATGATCCATCATCAGTAACAACATACACGGTGTAGTCCCAAATCCATCATA 534
Db     672 TDKKRDGKRGDGDRKDKGTGGDGDSKAKGDRAWAWRAKATRAAAADATAAATAAKATAWA 731

Qy     535 TGCACCTTCTTTTCTTCAATTTGGTCTTGTTTTTTTTTTTTTCATGATGTCATTGAATTAT 594
Db     732 TKTTTKKTTTTTTTKTTTTTTTTTTTTKKTTTARKADRDDDDKAKRDGDKDGDWDDADKDW 791

Qy     595 TCAAGAAGTCACTTCGAGCATAATGATTTTTCAAAA 630
Db     792 KAGRDRDDDTRKWGWKDRAGKKKKAKKKDHKTAWWW 827
```

RESULT 10
BM452144/c
LOCUS BM452144 1204 bp mRNA linear EST 05-FEB-2002
DEFINITION AGENCOURT_6386200 NIH_MGC_72 Homo sapiens cDNA clone IMAGE:5526492 5', mRNA sequence.
ACCESSION BM452144
VERSION BM452144.1 GI:18501184
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1204)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC/DCTD/DTP
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM12200 row: d column: 13
High quality sequence stop: 542.
FEATURES Location/Qualifiers
source 1. .1204
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5526492"
/clone_lib="NIH_MGC_72"
/tissue_type="melanotic melanoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: skin; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 2 kb. Library constructed by Life Technologies."
BASE COUNT 598 a 78 c 96 g 157 t 275 others
ORIGIN

Query Match 4.3%; Score 52.6; DB 13; Length 1204;
Best Local Similarity 28.8%; Pred. No. 4.6;
Matches 183; Conservative 0; Mismatches 448; Indels 4; Gaps 1;

```
Qy       1 TTTGAAATTTATGTATTTATCTATAGCATTAGAAACTATAAGAGTTGTTAGCTTCACTTG 60
Db    1108 TTTTTNTTNNNNTNTTTNTNTTTNTNNNTTTTTNNNNTTNTNNNTTTTTTTNTNNNTTTN 1049

Qy      61 GCTTACTGTTGTGCTCAAAGCAACTTCATCATCATACAGTATGGTTTTGATATGCTCTTC 120
Db    1048 TTTTTTTTTTNTTNTTTTNNTNCCCNNNNNTNNNAAAATTNNTTTTNNNNNCCCCNNNNN 989

Qy     121 CATTATCACTGAGCCTTATGATTATGTTTTACGAGCTTATAATATCACTGATGGTGATTC 180
Db     988 NNTNTTNNTTTTTTTTTNTTTTNTTTTTTTNTTTNNNTTTTTTTTTNTNTTTTTTTTT 929

Qy     181 AGTATTGTGATTATGTCCTTCGTTGATTATTCTGTTTCATACAAGTCGTGTAATTTGCTG 240
Db     928 TNNNTTNNNNNNTTTTTTTTTTNNTTTTTTTTTTTTTNNTNTTNNNTTTTTTTTTNTTT 869

Qy     241 TTTGTGACAGTACGATAGATCGACTCAACCTTCTGAGGTATTAGTTGAAGTTCATGTAAA 300
Db     868 TTNNTNTNNTTTTTTTTTNNNNTTTTNTTTTNNCCCCTTTTTTTTNNNNTTTTTNTNTT 809

Qy     301 TTAGCTTTGTTTATCATAGTAGCATTTGATTATTGATGCTCTGTAGCTAATGATAAGCCA 360
Db     808 TNNNNNNNNNNNTNNNNNNNTNTNTTTTNNNTTTTTTTNTNNNNNTNTTNNNNTTTTNNT 749

Qy     361 TTGGAGGGAAGCAAGCTTTC----TAAATGAATCTACGAATGGATGATAAAGTTCATGAA 416
Db     748 TTTTTTTTNNNNNNNNTTTNNTTTTTTNNTNTTTTTNNNNNTNNTTTNNNNTNTTTTNNN 689

Qy     417 TATTTTTGTTACTTCTGCAGTCAGATCATGAGTTATTGAGTCTATTGTTTTTTTAAGCCT 476
Db     688 TTTTTTTTTTTNTTTNTNTNNTTTTTTNNTTTNTTTTTNNTTNTNNTTTNNNNTTTNNNNT 629

Qy     477 GTTTCAGATGATCCATCATCAGTAACAACATACACGGTGTAGTCCCAAATCCATCATATG 536
Db     628 TTTNNNNNNNNTTNNNNNTNNNTNTNNNNNNNNNNNNNNNNNNNNNNNNNNNTNNNNTN 569

Qy     537 CACCTTCTTTTCTTCAATTTGGTCTTGTTTTTTTTTTTTCATGATGTCATTGAATTATTC 596
Db     568 NNNNNNNNNNNTTTTTTTNNNTNNNTNNTTTTNTTTTTTTTTTTTTTTTTTTTTTTTT 509

Qy     597 AAGAAGTCACTTCGAGCATAATGATTTTTCAAAAT 631
Db     508 TTTTTTTTTTTTTTTTTNNTATAGCTTTTGCTTATT 474
```

RESULT 11
CNS033GQ/c
LOCUS CNS033GQ 1135 bp DNA linear GSS 15-MAY-2000
DEFINITION Tetraodon nigroviridis genome survey sequence PUC-Ori end of clone 208P24 of library G from Tetraodon nigroviridis, genomic survey sequence.
ACCESSION AL226115
VERSION AL226115.1 GI:7885026
KEYWORDS GSS; genome survey sequence.
SOURCE Tetraodon nigroviridis.
ORGANISM Tetraodon nigroviridis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes; Tetraodontidae; Tetraodon.
REFERENCE 1 (bases 1 to 1135)
AUTHORS Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,

```
            Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
            Saurin,W. and Weissenbach,J.
  TITLE     Human gene number estimate provided by genome wide analysis using
            Tetraodon nigroviridis DNA sequence
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1135)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
            Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
            Weissenbach,J.
  TITLE     Charaterization and repeat analysis of the compact genome of the
            freshwater pufferfish Tetraodon nigroviridis
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 1135)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2000)
COMMENT     This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.
FEATURES             Location/Qualifiers
     source          1. .1135
                     /organism="Tetraodon nigroviridis"
                     /db_xref="taxon:99883"
                     /clone="208P24"
                     /clone_lib="G"
                     /note="Genoscope sequence ID : C0AG208DH12SP1-end :
                     PUC-Ori"
BASE COUNT      863 a     65 c     43 g    124 t     40 others
ORIGIN

  Query Match             4.3%;  Score 52.4;  DB 17;  Length 1135;
  Best Local Similarity   41.6%;  Pred. No. 5.2;
  Matches  224;  Conservative   13;  Mismatches  301;  Indels    0;  Gaps    0;

Qy     105 TTTTGATATGCTCTTCCATTATCACTGAGCCTTATGATTATGTTTTACGAGCTTATAATA 164
           ||||  | |  | ||  |||     |||:  || |  || | ||||   :| ||||| |
Db    1070 TTTTTTTTTTTTTTTGAATTTATTTTGARTTTTTTTTTTTTTTTTTTATWGWTTATATTT 1011

Qy     165 TCACTGATGGTGATTCAGTATTGTGATTATGTCCTTCGTTGATTATTCTGTTTCATACAA 224
           |   |  |   |  |     || | :|| | |  ||  ||  || || | |||  |
Db    1010 TTTTTTTTATAGTGTTTTATTTTTTTWTTTTTTTTTTTATTTTTTTTTTTTTTTTTTTTT 951

Qy     225 GTCGTGTAATTTGCTGTTTGTGACAGTACGATAGATCGACTCAACCTTCTGAGGTATTAG 284
           |  | |  || | |    || |   ||   |   |    |    || |      | |:
Db     950 TTTTTTTTTTTGGTTTAGAGTTATTTTATTTTTTGTTTTTTTTTGTTTTTTTTTATTGAR 891

Qy     285 TTGAAGTTCATGTAAATTAGCTTTGTTTATCATAGTAGCATTTGATTATTGATGCTCTGT 344
           || |  ||  | |   ||   | | ||| |  |  |    |||  || ||  |  | |:|
Db     890 TTTATTTTTTTATTTTTTTTTTNTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTRT 831

Qy     345 AGCTAATGATAAGCCATTGGAGGGAAGCAAGCTTTCTAAATGAATCTACGAATGGATGAT 404
           :|    | |        ||     :  ::  ||  | | |   | |   : | | |
Db     830 RGTGTGTTTTTGTWTTTTTTTTATTWATRRTKTTAATTATTTTTTTTTTTAWTTTGTTTTR 771

Qy     405 AAAGTTCATGAATATTTTTGTTACTTCTGCAGTCAGATCATGAGTTATTGAGTCTATTGT 464
           |  ||  |   | ||||| ||     || | |    |  |   || ||   | |||  |
Db     770 GAKTTTTTTTTTTTTTTTTTTTTTANATGTAATTTTTTTTTTTTTTTTTTTTATATATT 711

Qy     465 TTTTTTAAGCCTGTTTCAGATGATCCATCATCAGTAACAACATACACGGTGTAGTCCCAA 524
           ||||||      | ||    :  :|   |  |   |||  |  ||     | ||
Db     710 TTTTTTTTTTTTTTTATTAWATWTTTTTTTTTTTTTAAATATTTATTTTTTTTAAATTTTT 651

Qy     525 ATCCATCATATGCACCTTCTTTTCTTCAATTTGGTCTTGTTTTTTTTTTTTTCATGATGTC 584
           ||   |  |:      || |||| ||   |||  | || ||||||||||||  |  | |
Db     650 ATTTTTTTWATTTTTTTTTTTTTTTTTTTTTAATTTTWTTTTTTTTTTTTTTTTTTTATT 591

Qy     585 ATTGAATTATTCAAGAAGTCACTTCGAGCATAATGATTTTTCAAAATCCACCTTTGTT 642
           |   || ||      |   ||     |  ||  |||||     |  |  ||| ||
Db     590 TTATTTTTTTTTTTTTTTTTTTTTTATTTAATATTTTTTTTTTTTTTTTAATTTTTTT 533


RESULT 12
```

```
AZ548700
LOCUS       AZ548700                872 bp    DNA     linear   GSS 14-NOV-2000
DEFINITION  ENTGE67TF Entamoeba histolytica Sheared DNA Entamoeba histolytica
            genomic, DNA sequence.
ACCESSION   AZ548700
VERSION     AZ548700.1  GI:11172565
KEYWORDS    GSS.
SOURCE      Entamoeba histolytica.
  ORGANISM  Entamoeba histolytica
            Eukaryota; Entamoebidae; Entamoeba.
REFERENCE   1  (bases 1 to 872)
  AUTHORS   Loftus,B., Van Aken,S. and Fraser,C.
  TITLE     Determination of clone end sequences from Entamoeba histolytica
            HM1:IMSS sheared DNA library
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Brendan J Loftus
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0208
            Fax: 301 838 3543
            Email: bjloftus@tigr.org
            Clones are derived from the Entamoeba histolytica HM1:IMSS sheared
            DNA library
            Seq primer: M13-Forward
            Class: shotgun
            High quality sequence start: 33
            High quality sequence stop: 800.
FEATURES             Location/Qualifiers
     source          1. .872
                     /organism="Entamoeba histolytica"
                     /strain="HM1:IMSS"
                     /db_xref="taxon:5759"
                     /clone_lib="Entamoeba histolytica Sheared DNA"
                     /note="Vector: pHOS1; Site_1: Bst I; Constructed at The
                     Institute for Genomic Research (TIGR), Rockville, MD.
                     Genomic DNA isolated from broth cultures of E. histolytica
                     using a  method described by Clark and Diamond (Clark,
                     C.G., and Diamond, L.S.  (1993) Entamoeba histolytica: a
                     method for isolate identification.  Exp. Parasitol.
                     77:450.). The DNA was mechanically sheared to give a
                     tight size distribution (~2 kb). The v + i method used for
                     the library  construction is described in detail in Smith,
                     H.O. and Venter, J.C.  (Making small insert libraries for
                     whole genome shotgun sequencing  projects. In Genome
                     Sequencing: A Practical Approach, eds. M. Vaudin  and B.
                     Barell, Oxford University Press, 1999)."
BASE COUNT      394 a     57 c     59 g    362 t
ORIGIN

  Query Match             4.3%;  Score 52;  DB 17;  Length 872;
  Best Local Similarity   53.4%;  Pred. No. 6.9;
  Matches  109;  Conservative    0;  Mismatches   95;  Indels    0;  Gaps    0;

Qy     378 TTCTAAATGAATCTACGAATGGATGATAAAGTTCATGAATATTTTTGTTACTTCTGCAGT 437
           ||  |||| |   || ||||| ||  ||||  | ||| | | || | |      | | |
Db     669 TTAAAAATTAGAATAAGAATGTATTTTAAAAATTATGTAAAATTATTTATTCATTTTAAT 728

Qy     438 CAGATCATGAGTTATTGAGTCTATTGTTTTTTTAAGCCTGTTTCAGATGATCCATCATCA 497
           |||||  |  |    | | |||| || ||| | |  |||||   |      | || |
Db     729 TTAATCATAACATTACAAATTTATTTTTATTTAATGAATGTTTGTATTTCAAACTAATAA 788

Qy     498 GTAACAACATACACGGTGTAGTCCCAAATCCATCATATGCACCTTCTTTTCTTCAATTTG 557
           || |||| ||  |   |  |     ||| || || |    | | | || || ||| |
Db     789 ACAAAAACAAACTGGTCATGTTAAATTATCGATAATTTAATTCGTTTATTTTTAAATATT 848

Qy     558 GTCTTGTTTTTTTTTTTTCATGAT 581
           |||  ||  |||| |||||  ||
Db     849 TTCTATTTCCTTTTATTTCAAAAT 872


RESULT 13
CNS000B8
```

```
LOCUS       CNS000B8                1101 bp    DNA     linear   GSS 03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence TET3 end of BAC #
            BACR01A24 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL063632
VERSION     AL063632.1  GI:4938680
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .1101
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone="BACR01A24"
                     /clone_lib="RPCI-98"
                     /note="end : TET3"
BASE COUNT      228 a    114 c    110 g    512 t    137 others
ORIGIN

  Query Match             4.3%;  Score 52;  DB 17;  Length 1101;
  Best Local Similarity   36.5%;  Pred. No. 6.2;
  Matches  148;  Conservative   50;  Mismatches  208;  Indels    0;  Gaps    0;

Qy    190 ATTATGTCCTTCGTTGATTATTCTGTTTCATACAAGTCGTGTAATTTGCTGTTTGTGACA 249
          | ||| |  ||   : :||:|| |  :| :     |:  | ||||||  | |:| | :
Db    683 AATATTTTTTTKTGKTWTTDTTTTTGWTTWATTTKGDAATWTAATTTTTTWTWTATWWWT 742

Qy    250 GTACGATAGATCGACTCAACCTTCTGAGGTATTAGTTGAAGTTCATGTAAATTAGCTTTG 309
           |:  |:|  |  : | |   |       :  || : | |  || || ||: ||   : |
Db    743 WTWTWAWAWTTTWWTTTATTTTRAATTWRATTTTKGTAATTTTWATTTADTTTTTTWATT 802

Qy    310 TTTATCATAGTAGCATTTGATTATTGATGCTCTGTAGCTAATGATAAGCCATTGGAGGGA 369
          |:|   :|  |:   :::  |::||  |  | | |   |  | ||      ||  :  |
Db    803 TWTTATWTTWTWTATWWWTTTWWTTTTTWATTTATTTTTTTTWATTTTTTTTTTAWTAAA 862

Qy    370 AGCAAGCTTTCTAAATGAATCTACGAATGGATGATAAAGTTCATGAATATTTTTGTTACT 429
          |  ||   |   |  |: |:|  |     |  ||  | :| ::    | ||| |:| |:| :
Db    863 AAAAAATTAAATTTAWWAWTTAAAWTTTTWATTTTTWATWWTTAAATTATATWTTTWATW 922

Qy    430 TCTGCAGTCAGATCATGAGTTATTGAGTCTATTGTTTTTTTAAGCCTGTTTCAGATGATC 489
          : |  | :   ||  :   || ||   | | || ||||||: |   :  |:   |   |
Db    923 WTTAWAAWTTTATTTWTTTTTTTTTTATTTTTTTTTTTTTTWTAWATWTATWTTTAAWTTT 982

Qy    490 CATCATCAGTAACAACATACACGGTGTAGTCCCAAATCCATCATATGCACCTTCTTTTCT 549
          | |  |      :   : |   |     |   || |  ||  | |  |   : :: : |
Db    983 ATTAAAWATATTATWTTAWWATTTTTATATTAWAATTTTATTTTTTTTTATTAWTWWAWAT 1042

Qy    550 TCAATTTGGTCTTGTTTTTTTTTTTTTCATGATGTCATTGAATTATT 595
          :  :|||    || ||||||||||:: |  || | |||  :||:||
```

```
Db    1043 WATWTTTATAATTTTTTTTTTTTTWWAAATATTTAATTTTWTTWTT 1088


RESULT 14
CNS020K7/c
LOCUS       CNS020K7                1092 bp    DNA     linear   GSS 12-MAY-2000
DEFINITION  Tetraodon nigroviridis genome survey sequence T7 end of clone
            222L11 of library G from Tetraodon nigroviridis, genomic survey
            sequence.
ACCESSION   AL175696
VERSION     AL175696.1  GI:7813753
KEYWORDS    GSS; genome survey sequence.
SOURCE      Tetraodon nigroviridis.
  ORGANISM  Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetraodontidae; Tetraodon.
REFERENCE   1  (bases 1 to 1092)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
            Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
            Saurin,W. and Weissenbach,J.
  TITLE     Human gene number estimate provided by genome wide analysis using
            Tetraodon nigroviridis DNA sequence
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1092)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
            Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
            Weissenbach,J.
  TITLE     Charaterization and repeat analysis of the compact genome of the
            freshwater pufferfish Tetraodon nigroviridis
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 1092)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2000)
COMMENT     This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.
FEATURES             Location/Qualifiers
     source          1. .1092
                     /organism="Tetraodon nigroviridis"
                     /db_xref="taxon:99883"
                     /clone="222L11"
                     /clone_lib="G"
                     /note="Genoscope sequence ID : C0AG222CF06LP1-end : T7"
BASE COUNT      383 a    169 c    165 g    262 t    113 others
ORIGIN

  Query Match             4.2%;  Score 50.8;  DB 17;  Length 1092;
  Best Local Similarity   37.1%;  Pred. No. 10;
  Matches  182;  Conservative   63;  Mismatches  238;  Indels    7;  Gaps    2;

Qy    106 TTTGATATGCTCTTCCATTATCACTGAGCCTTATGATTATGTTTTACGAGCTTATAATAT 165
          |||  | |  | ||   || |   | |   ||    |||| ||||       |||  ||||
Db   1087 TTTTTTTTTTTTTTTTTTTTTATTTAWTTTTTAATTTATTTTTTTTTTTTTTAATATAT 1028

Qy    166 CACTGATGGTGATTCAGTATTGTGATTATGTCCTTCGTTGATTATTCTGTTTCATACAAG 225
          | |  |      |:       | | ||| | |  ||  || |  | |    ||  ::
Db   1027 AAATATTTTATTTWTTTATATTTTATTTTTTTTTTTTTTTTAAAAATAAATTAAWWTTTTN 968

Qy    226 TCGTGTAATTTGCTGTTTGTGACAGTACGATAGATCGACTCAACCTTCTGAGGTATTAGT 285
          :  | || |||  | ||| |   |  |  |:: |   |   ||  :  | |  ::|:|
Db    967 WTTTATATTTTTTTTTTTTWTTAWAAAAAWWAAAAAAAAAAAAAAWATTTAWAWWTWA-- 910

Qy    286 TGAAGTTCATGTAAATTAGCTTTGTTTATCATAGTAGCATTTGATTATTGATGCTCTGTA 345
          : ||    |:  :  :|   || ||| |  |  ||    |:: ||| :| |   : | ||
Db    909 WAAAAAAAAWWAWTTWTTTTTTAATTTTTTTTTTTAWATTWWAATTTWTAAAWTWTTTTA 850

Qy    346 GCTAATGATAAGCCATTGGAGGGAAGCAAGCTTTCTAAATGAATCTACGAATGGATGATA 405
          :|||  ||     ||  |   |:  ||   || |:  | ||   :  ::|   : :||
Db    849 AAWAATTATTTWWTTTTTTAWTAAWTAAAWTNTTTTWTTTAAAATAWAAWWTWATWWWTA 790
```

```
Qy     406 AAGTTCATGAATATTTTTGTTACTTCTGCAGTCAGATCATGAGTTATTGAGTCTATTGTT 465
           || :| |: ::: ||:|| |:: :: :  |   | :  :| : ||:||   | | :: ||
Db     789 AAAWTAAWAWWWTTTWTTTTWWWWWWWTWAAAAAAWAAWTWWWTTWTTWTTTTTTWWTTT 730

Qy     466 TTTTTAAGCCTGTTTCAGATGATCCATCATCAGTAACAACATACACGGTGTAGTCCCAAA 525
           : ||:|:   | |||    | .|  |  |  |  || || | |     | |  |   :::
Db     729 WATTWAWTTTTTTTTTTTTTTTTTAAAAAAAAAAAAAAAAAAAAATTTTTTTTTTAAWWWW 670

Qy     526 TCCATCATATGCACCTTCTTTTCTTCAATTTGGTCTTGTTTTTTTTTTTTTCATGATGTCA 585
           :  :| :| |     || |:|     || ::  | || ||||||::::|| .|  ||   :
Db     669 WWWWTTWTTT-----TTTTWTAAAAAAAAWWTTTTTTTTTTTTTTWWWWTTAAAWATAAAW 615

Qy     586 TTGAATTATT 595
           || :|  |::
Db     614 TTTWAAAAWD 605


RESULT 15
AL546530/c
LOCUS       AL546530                604 bp    mRNA    linear   EST 16-FEB-2001
DEFINITION  AL546530 LTI_NFL006_PL2 Homo sapiens cDNA clone CS0DI030YL05 5
            prime, mRNA sequence.
ACCESSION   AL546530
VERSION     AL546530.1  GI:12879736
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 604)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
FEATURES             Location/Qualifiers
     source          1. .604
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="CS0DI030YL05"
                     /clone_lib="LTI_NFL006_PL2"
                     /tissue_type="placenta"
                     /note="Vector: pCMVSPORT 6; Site_1: NotI; 1st strand cDNA
                     was primed with a NotI-oligo(dT) primer.  Five prime end
                     enriched, double-stranded cDNA was digested with Not I and
                     cloned into the Not I and Eco RV sites of the pCMVSPORT 6
                     vector. Library was normalized. Library was constructed by
                     Life Technologies. Contact : Feng Liang Life Technologies,
                     a division of Invitrogen 9800 Medical Center Drive
                     Rockville, Maryland 20850, USA Fax : (1) 301 610 8371
                     Email : fliang@lifetech.com URL :
                     http://fulllength.invitrogen.com"
BASE COUNT      163 a     88 c    103 g    212 t     38 others
ORIGIN

  Query Match             4.1%;  Score 50.4;  DB 9;  Length 604;
  Best Local Similarity   43.6%;  Pred. No. 16;
  Matches  115;  Conservative   21;  Mismatches  128;  Indels    0;  Gaps    0;

Qy     886 AGAAAAGAAAGACCAAAACAATGAAAACTTTACGAAAAATCAAAAAGTTGAAGGACTTTA 945
           | |||| ||| |  |||| ||  ||||    |  |||||  ||||| : |||| | || |
Db     564 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAWAGAAGAAATTAA 505

Qy     946 ACGTCGAGATCTCTCGTAGAAAACCTCTTTTGTAAGGTTGCATACAATACTTTTTTTTCA 1005
           |     | |    |   | ||||  | : :: |::  ::  :||  ::| ||||||||
Db     504 AAAAAAAAAAAATTRWAAAAAAAWWTTWAWDTTWWTWKWTTWTAWTWWAWTTTTTTTTTT 445

Qy    1006 GACTTTACTTATGGTATTATACTGAATATGTTATTGCTGTTATAGTAGTTGAGTGACGTT 1065
           |  ||:  || |  | :: :: |   : | || ||  | .|| |  |   |   |    ||
Db     444 GGGTTYTTTTTTTTTTTWWKWWWTTTKKKTTTTTTTTTTTTTTTTTCTCTTTTTTKKTTT 385

Qy    1066 TGAGGGAATTTCTAGTCCGTTAATCTTGTACTCAGTGTGTCTACTTTTCAAAAAAGTCAG 1125
           |       ||| :  ::: ||  | || |  |   | | | | |       | || |
Db     384 TTTTTTTTTTTTYYYYYYYTTTTTTTTTTTTTTTTTTTTTTTCAAAAATGACAAATTCT 325

Qy    1126 TTTTTCAGTCTCTAAAACACATTT 1149
           || || |       |||| ||| |
Db     324 TTATTTAAATCAACAAACTCATCT 301


Search completed: May 24, 2003, 08:04:04
Job time : 1718 secs
```

GenCore version 5.1.5

OM nucleic - nucleic search, using sw model

Run on: May 24, 2003, 07:52:14 ; Search time 185 Seconds
(without alignments)
8686.500 Million cell updates/sec

Title: US-09-701-926B-1
Perfect score: 1217
Sequence: 1 tttgaaatttatgtatttat.....gtcaacacaacacaacaaca 1217

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 828747 seqs, 660231138 residues

Total number of hits satisfying chosen parameters: 1657494

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	44.8	3.7	5610	9	US-10-239-676-169	Sequence 169, App
2	44.4	3.6	5689	9	US-10-239-676-90	Sequence 90, Appl
3	44.4	3.6	8842	9	US-10-239-676-72	Sequence 72, Appl
4	43.2	3.5	12968	9	US-10-239-676-202	Sequence 202, App
5	42.8	3.5	7903	9	US-10-239-676-110	Sequence 110, App
c 6	42.4	3.5	3111	10	US-09-815-242-4518	Sequence 4518, Ap
c 7	42.4	3.5	3198	10	US-09-815-242-8519	Sequence 8519, Ap
8	40.4	3.3	6282	9	US-10-239-676-127	Sequence 127, App
9	40.2	3.3	12405	9	US-10-239-676-35	Sequence 35, Appl
10	40	3.3	15732	9	US-10-239-676-95	Sequence 95, Appl
11	39.6	3.3	426	10	US-09-960-352-8406	Sequence 8406, Ap
c 12	39.6	3.3	2093	9	US-10-005-530-42	Sequence 42, Appl
c 13	39.6	3.3	640681	10	US-09-790-988-1	Sequence 1, Appli
14	39.2	3.2	9293	9	US-10-239-676-26	Sequence 26, Appl
15	39	3.2	6306	9	US-10-239-676-129	Sequence 129, App
16	38.8	3.2	414	10	US-09-960-352-6528	Sequence 6528, Ap
17	38.8	3.2	7657	9	US-10-239-676-185	Sequence 185, App
c 18	38.6	3.2	2000	9	US-09-938-842A-3858	Sequence 3858, Ap
19	38.4	3.2	530	9	US-09-796-692-8670	Sequence 8670, Ap
20	38.4	3.2	530	9	US-10-040-862-8670	Sequence 8670, Ap
21	38.4	3.2	6327	9	US-10-239-676-151	Sequence 151, App
22	38.4	3.2	9539	9	US-10-239-676-52	Sequence 52, Appl
c 23	38.4	3.2	513509	9	US-09-754-853A-4	Sequence 4, Appli
c 24	38.2	3.1	1813	10	US-09-880-578-3	Sequence 3, Appli
c 25	38.2	3.1	32082	9	US-09-764-891-9679	Sequence 9679, Ap
26	38	3.1	355	10	US-09-960-352-14757	Sequence 14757, A
27	38	3.1	45862	9	US-10-216-355-3	Sequence 3, Appli
28	37.8	3.1	1819	9	US-10-125-237-87	Sequence 87, Appl
29	37.8	3.1	1819	9	US-10-105-891-87	Sequence 87, Appl
30	37.8	3.1	9515	9	US-10-239-676-160	Sequence 160, App
c 31	37.8	3.1	11464	12	US-10-100-057-17	Sequence 17, Appl
c 32	37.6	3.1	11036	9	US-10-239-676-117	Sequence 117, App
c 33	37.6	3.1	1503841	9	US-09-946-807-1	Sequence 1, Appli
c 34	37.6	3.1	1503841	10	US-09-795-668-1	Sequence 1, Appli
c 35	37.6	3.1	1503841	10	US-09-795-686-1	Sequence 1, Appli
c 36	37.2	3.1	570	9	US-09-918-995-12139	Sequence 12139, A
37	37.2	3.1	6069	9	US-10-239-676-172	Sequence 172, App
38	37.2	3.1	7657	9	US-10-239-676-186	Sequence 186, App
39	37.2	3.1	1691139	9	US-10-067-514-1	Sequence 1, Appli
40	37	3.0	286	10	US-09-960-352-13342	Sequence 13342, A
41	37	3.0	2000	9	US-09-938-842A-3028	Sequence 3028, Ap
c 42	37	3.0	2000	9	US-09-938-842A-4258	Sequence 4258, Ap
43	37	3.0	6030	9	US-10-239-676-164	Sequence 164, App
44	37	3.0	70768	9	US-10-135-322-13	Sequence 13, Appl
45	36.8	3.0	2000	9	US-09-938-842A-3483	Sequence 3483, Ap

ALIGNMENTS

RESULT 1
US-10-239-676-169
; Sequence 169, Application US/10239676
; Publication No. US20030082609A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with Gene Regulation
; FILE REFERENCE: 5013.1003
; CURRENT APPLICATION NUMBER: US/10/239,676
; CURRENT FILING DATE: 2002-09-24
; PRIOR APPLICATION NUMBER: PCT/EP01/03968
; DE 10019058.8
; DE 10019173.8
; DE 10032529.7
; DE 10043826.1
; PRIOR FILING DATE: 2001-04-06
; 2000-04-06
; 2000-04-07
; 2000-06-30
; 2000-09-01
; NUMBER OF SEQ ID NOS: 228
; SEQ ID NO 169
; LENGTH: 5610
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-239-676-169

Query Match 3.7%; Score 44.8; DB 9; Length 5610;
Best Local Similarity 52.1%; Pred. No. 1.4;
Matches 100; Conservative 0; Mismatches 92; Indels 0; Gaps 0;

```
Qy    279 TATTAGTTGAAGTTCATGTAAATTAGCTTTGTTTATCATAGTAGCATTTGATTATTGATG 338
          |||  ||| || |  |||| ||||  ||| ||| |  ||    |  | | |||| ||
Db   1084 TATATTTTGGAGATGTTGTATATTATGTTTTTTTTTTTAAGAGTTACGTAAGTATTTATA 1143

Qy    339 CTCTGTAGCTAATGATAAGCCATTGGAGGGAAGCAAGCTTTCTAAATGAATCTACGAATG 398
          | || |  ||| |||  ||| | |  ||      |||     | ||| ||  ||||
Db   1144 TATGGAAGATTTTGAGAAGGTATTTGGGAAAAAATTTGTTTAATTGTAAATATAATAATG 1203
```

```
Qy     399 GATGATAAAGTTCATGAATATTTTTGTTACTTCTGCAGTCAGATCATGAGTTATTGAGTC 458
           |||     ||   || |  ||||| ||   |      |      ||   ||||||  |
Db    1204 TATGTATTTGTATGTGGAGTTTTTTTTTTAGTGGATTATTTATAAATCGTTTATTGTTTT 1263

Qy     459 TATTGTTTTTTT 470
           || ||| | |||
Db    1264 TAATGTATATTT 1275


RESULT 2
US-10-239-676-90
; Sequence 90, Application US/10239676
; Publication No. US20030082609A1
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with Gene Regulation
;  FILE REFERENCE: 5013.1003
;  CURRENT APPLICATION NUMBER: US/10/239,676
;  CURRENT FILING DATE:  2002-09-24
;  PRIOR APPLICATION NUMBER: PCT/EP01/03968
;    DE 10019058.8
;    DE 10019173.8
;    DE 10032529.7
;    DE 10043826.1
;  PRIOR FILING DATE: 2001-04-06
;    2000-04-06
;    2000-04-07
;    2000-06-30
;    2000-09-01
;  NUMBER OF SEQ ID NOS: 228
; SEQ ID NO 90
;   LENGTH: 5689
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (1818, 1930..1931, 1934..1935, 1947..1948, 1966)
US-10-239-676-90

  Query Match             3.6%;  Score 44.4;  DB 9;  Length 5689;
  Best Local Similarity   43.9%;  Pred. No. 1.8;
  Matches  215;  Conservative    0;  Mismatches  274;  Indels    1;  Gaps    1;

Qy     107 TTGATATGCTCTTCCATTATCACTGAGCCTTATGATTATGTTTTACGAGCTTATAATATC 166
           || ||||  | ||   || | |  ||   |      || |||||      | ||      ||
Db    1600 TTAATATTTTATTTATTTTTGATAGATTATATGTTTTTTGTTTGTTTTGTTTTATTAATT 1659

Qy     167 ACTGATGGTGATTCAGTATTGTGATTATGTCCTTCGTTGATTATTCTGTTTCATACAAGT 226
               |||  ||||| | |||   |  |   ||   | || | ||   |   | |
Db    1660 GAATTAATATATTGCGTATTTTTATTTGTTATTATTTTTCGTTTTTTTTTATTTTTTAAT 1719

Qy     227 CGTGTAATTTGCTGTTTGTGACAGTACGATAGATCGACTCAACCTTCTGAGGTATTAGTT 286
           | | ||||  | ||| | |   ||   || ||     |       | | |  ||||  ||
Db    1720 ATTTTTATTTTTTTTTTTATTATTTTATTTTATATATTATTTTTTATTTTATATATTTTTT 1779

Qy     287 GAAGTTCATGTAAATTAGCTTTGTTTATCATAGTAGCATTTGATTATTGATGCTCTGTAG 346
           | |   | |  |||   ||| | | | ||| |      | ||||||     |   ||
Db    1780 ATATTATTTTTTTATT-TATTTTTATTTTATAATTTTTANTAATTATTTTATTTTAATAT 1838

Qy     347 CTAATGATAAGCCATTGGAGGGAAGCAAGCTTTCTAAATGAATCTACGAATGGATGATAA 406
           ||   | |     |  |  ||      ||| |    |  | ||     |  || | |
Db    1839 ATATAAAATAAATTATTTAATTAATATTTATTTTTTTTTATTTATATTTTTTTATTAATA 1898

Qy     407 AGTTCATGAATATTTTTGTTACTTCTGCAGTCAGATCATGAGTTATTGAGTCTATTGTTT 466
           | ||  | |   | ||  |   || |    |   |   |   || |   |  |   ||
Db    1899 AATTTTTTATATTATTATTATTTTTTTTTTATNNTANNTTTTTTTTTAANNTTATTAAATT 1958

Qy     467 TTTTAAGCCTGTTTCAGATGATCCATCATCAGTAACAACATACACGGTGTAGTCCCAAAT 526
```

```
           |||| |   | ||| |  |  |     || |  || || | | |   | ||     ||||
Db    1959 TTTTTATNTTTTTTTAATTATTTTTAAATAATAAAAAAAAAAAAAATATATAAAAATAAAT 2018

Qy     527 CCATCATATGCACCTTCTTTTCTTCAATTTGGTCTTGTTTTTTTTTTTTTCATGATGTCAT 586
             ||  | |     |  | || ||  ||   | || ||||||||        |  |  |
Db    2019 TTATTTTTTTTTTTTATTGTTTTTTTTTTATATTTTTTTTTTTTTTTCGGAGGAAAATTTT 2078

Qy     587 TGAATTATTC 596
           |  |||||||
Db    2079 TTGATTATTC 2088


RESULT 3
US-10-239-676-72
; Sequence 72, Application US/10239676
; Publication No. US20030082609A1
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with Gene Regulation
;  FILE REFERENCE: 5013.1003
;  CURRENT APPLICATION NUMBER: US/10/239,676
;  CURRENT FILING DATE:  2002-09-24
;  PRIOR APPLICATION NUMBER: PCT/EP01/03968
;    DE 10019058.8
;    DE 10019173.8
;    DE 10032529.7
;    DE 10043826.1
;  PRIOR FILING DATE: 2001-04-06
;    2000-04-06
;    2000-04-07
;    2000-06-30
;    2000-09-01
;  NUMBER OF SEQ ID NOS: 228
; SEQ ID NO 72
;   LENGTH: 8842
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-239-676-72

  Query Match             3.6%;  Score 44.4;  DB 9;  Length 8842;
  Best Local Similarity   47.9%;  Pred. No. 2.3;
  Matches  160;  Conservative    0;  Mismatches  171;  Indels    3;  Gaps    1;

Qy      12 TGTATTTATCTATAGCATTAGAAACTATAAGAGTTGTTAGCTTCACTTGGCTTACTGTTG 71
           | | ||||| | |     |    |  |       ||| ||   ||   ||| ||   |
Db    6699 TTTTTTTATTTTTTTTTTGTTGATTTTATTTTGTTATTTTGTTTTTTTGTTTTTAGGGGT 6758

Qy      72 TGCTCAAAGCAACTTCATCATCATACAGTATGGTTTTGATATGCTCTTCCATTATCACTG 131
           |  | |  |  | ||     |   ||   | |||| || | | | | |     |    |
Db    6759 TTTTTATTGGTAGTTATATAGGTTAATTTTTGGTATTTAAAAGTTATATTTATTAATTTT 6818

Qy     132 AGCCTTATGATTATGTTTTACGAGCTTATAATATCACTGATGGTGATTCAGTATTGTGAT 191
           |   ||   ||  | |  || ||  || |  | | |  ||||||  || | ||| ||
Db    6819 ATTTTTTATATATTTTCGTAGGAATTTTTGTTTTTTATGGATGGTAGTTTATTATAGTAGG 6878

Qy     192 TATGTCCTTCGTTGATTATTCTGTTTCATACAAGTCGTGTAATTTGCTGTTTGTGACAGT 251
           | ||      |||   | ||   ||  ||  ||||  | | ||||  |   ||||  |||
Db    6879 TTTG---GGAGTTAGGTTTTAGTTTATATTTAAGTTTTTTGATTTATTAAATGTGTGAGT 6935

Qy     252 ACGATAGATCGACTCAACCTTCTGAGGTATTAGTTGAAGTTCATGTAAATTAGCTTTGTT 311
           ||   |  || | ||  ||  || || ||||||         ||||  | |  |||  |
Db    6936 TTGAGTAAATGATTTAATTTTTCGAAGTTTTAGTTTTTTAATTTGTATAATGGTATTGAT 6995

Qy     312 TATCATAGTAGCATTTGATTATTGATGCTCTGTA 345
           |  ||| |||  |    ||  ||||  | | ||
Db    6996 GGTAATATTAGGTTAAATTTTGTGATATTATTTA 7029


RESULT 4
```

```
US-10-239-676-202
; Sequence 202, Application US/10239676
; Publication No. US20030082609A1
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with Gene Regulation
;  FILE REFERENCE: 5013.1003
;  CURRENT APPLICATION NUMBER: US/10/239,676
;  CURRENT FILING DATE:  2002-09-24
;  PRIOR APPLICATION NUMBER: PCT/EP01/03968
;    DE 10019058.8
;    DE 10019173.8
;    DE 10032529.7
;    DE 10043826.1
;  PRIOR FILING DATE: 2001-04-06
;    2000-04-06
;    2000-04-07
;    2000-06-30
;    2000-09-01
;  NUMBER OF SEQ ID NOS: 228
; SEQ ID NO 202
;   LENGTH: 12968
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-239-676-202

  Query Match             3.5%;  Score 43.2;  DB 9;  Length 12968;
  Best Local Similarity   49.3%;  Pred. No. 5.4;
  Matches  140;  Conservative    0;  Mismatches  143;  Indels    1;  Gaps    1;

Qy     183 TATTGTGATTATGTCCTTCGTTGATTATTCTGTTTCATACAAGTCGTGTAATTTGCTGTT 242
           ||| || | | |||  |   || ||| || | ||  | |    | ||||| |||     |
Db   11725 TATGGTTAATGTGTGATGATTTTATTTTTATTTTATAAATTTATTGTGTATTTTATAAGT 11784

Qy     243 TGTGACAGTACGATAGATCGACTCAACCTTCTGAGGTATTAGTTGAAGTTCATGTAAATT 302
           || |    | ||||  | |    |  || |     |||     | ||| | ||||||
Db   11785 AGTAAGTTGAGGATATTTTAAAATTAATTTTTTTAAAATTTAGGAATGTTTA-GTAAATA 11843

Qy     303 AGCTTTGTTTATCATAGTAGCATTTGATTATTGATGCTCTGTAGCTAATGATAAGCCATT 362
             | | ||||| | | |    | | |||||||  |    | ||| |||| |   |   |
Db   11844 TTTTATTTTTATAAAAATTTTGTATTATTATTGTTATAATATAGGTAATAAGTTGGGTTG 11903

Qy     363 GGAGGGAAGCAAGCTTTCTAAATGAATCTACGAATGGATGATAAAGTTCATGAATATTTT 422
           ||||||||  |  |||     ||  | ||   |    |    | ||   | |   | ||
Db   11904 TGAGGGAAGTTAAGTTTTGTTTTGTTTTTAGAGAGTTGTATAGAGGTATTTAATAGTATT 11963

Qy     423 TGTTACTTCTGCAGTCAGATCATGAGTTATTGAGTCTATTGTTT 466
           |    |  |   | | |   ||    ||||||  |||| |||
Db   11964 AGAATGTATTTATGGTAAACGTTGTAGAATTGAGGGTATTATTT 12007


RESULT 5
US-10-239-676-110
; Sequence 110, Application US/10239676
; Publication No. US20030082609A1
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with Gene Regulation
;  FILE REFERENCE: 5013.1003
;  CURRENT APPLICATION NUMBER: US/10/239,676
;  CURRENT FILING DATE:  2002-09-24
;  PRIOR APPLICATION NUMBER: PCT/EP01/03968
;    DE 10019058.8
;    DE 10019173.8
;    DE 10032529.7
;    DE 10043826.1
;  PRIOR FILING DATE: 2001-04-06
;    2000-04-06
;    2000-04-07
;    2000-06-30
;    2000-09-01
;  NUMBER OF SEQ ID NOS: 228
; SEQ ID NO 110
;   LENGTH: 7903
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-239-676-110

  Query Match             3.5%;  Score 42.8;  DB 9;  Length 7903;
  Best Local Similarity   50.5%;  Pred. No. 5.2;
  Matches  104;  Conservative    0;  Mismatches  102;  Indels    0;  Gaps    0;

Qy     974 TTTGTAAGGTTGCATACAATACTTTTTTTTCAGACTTTACTTATGGTATTATACTGAATA 1033
           ||| ||| | |  ||     |  || | |     | || | || |  | || |  |  ||
Db    1479 TTTTTAAAGATTTATGGTTTTGTTATATAAAGTATTTGAATTTTTTTTTTGTTTTTTATT 1538

Qy    1034 TGTTATTGCTGTTATAGTAGTTGAGTGACGTTTGAGGGAATTTCTAGTCCGTTAATCTTG 1093
           ||||  || | || |  | |||   |     |||  |    ||| ||||   ||  | ||
Db    1539 TGTTTGTGGTTTTTTTTTTGTTTTTTTTTTTTTTTTGTTTTTTTTTAGTTTTTTTTTTTT 1598

Qy    1094 TACTCAGTGTGTCTACTTTTCAAAAAAGTCAGTTTTTCAGTCTCTAAAACACATTTAAAT 1153
           ||   ||  | | |  |||| | |||  | ||  | |   | |  ||     || | | |
Db    1599 TATCGAGGATTTTTGTTTTTTAGAAATTTTAGGGTATTTATTTAGAATTTTTATATTAGT 1658

Qy    1154 AAGAGTTTCTTTGCCCATCTTTTGTT 1179
           |||| ||  || |      |||| ||
Db    1659 AAGAATTAATTCGTGTTATTTTTTTT 1684


RESULT 6
US-09-815-242-4518/c
; Sequence 4518, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
;  APPLICANT: Haselbeck, Robert
;  APPLICANT:  Ohlsen, Kari L.
;  APPLICANT:  Zyskind, Judith W.
;  APPLICANT:  Wall, Daniel
;  APPLICANT:  Trawick, John D.
;  APPLICANT:  Carr, Grant J.
;  APPLICANT:  Yamamoto, Robert T.
;  APPLICANT:  Xu, H. Howard
;  TITLE OF INVENTION: Identification of Essential Genes in
;  TITLE OF INVENTION:  Prokaryotes
;  FILE REFERENCE: ELITRA.011A
;  CURRENT APPLICATION NUMBER: US/09/815,242
;  CURRENT FILING DATE:  2001-03-21
;  PRIOR APPLICATION NUMBER: 60/191,078
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: 60/206,848
;  PRIOR FILING DATE: 2000-05-23
;  PRIOR APPLICATION NUMBER: 60/207,727
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: 60/242,578
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/253,625
;  PRIOR FILING DATE: 2000-11-27
;  PRIOR APPLICATION NUMBER: 60/257,931
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: 60/269,308
;  PRIOR FILING DATE: 2001-02-16
;  NUMBER OF SEQ ID NOS: 14110
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 4518
;   LENGTH: 3111
;   TYPE: DNA
;   ORGANISM: Staphylococcus aureus
US-09-815-242-4518
```

Query Match 3.5%; Score 42.4; DB 10; Length 3111;
Best Local Similarity 53.7%; Pred. No. 4;
Matches 88; Conservative 0; Mismatches 76; Indels 0; Gaps 0;

```
Qy    381 TAAATGAATCTACGAATGGATGATAAAGTTCATGAATATTTTTGTTACTTCTGCAGTCAG 440
          ||| || |||    |||   |  | |  ||||| || | |||     ||       |||
Db   2599 TAAGTGCATCAGGCAATTCTTCTTTATCTTCATACATCTGTTTTGGCGTTAAAATATCAA 2540

Qy    441 ATCATGAGTTATTGAGTCTATTGTTTTTTTAAGCCTGTTTCAGATGATCCATCATCAGTA 500
          || |  |   ||   ||  | |||||   | | | |   ||  ||||| ||     | |
Db   2539 ATAAAAAACCATCTTGTTCAATGTTTAAATCACCATCCAACACTTGATCAATAGCTTGCA 2480

Qy    501 ACAACATACACGGTGTAGTCCCAAATCCATCATATGCACCTTCT 544
          |||||  |  |  ||||  | |||||  ||||| | || ||  ||
Db   2479 ACAACGTTGAACGTGTTTTACCAAAAGCATCAAACGCTCCCACT 2436
```

RESULT 7
US-09-815-242-8519/c
; Sequence 8519, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 8519
; LENGTH: 3198
; TYPE: DNA
; ORGANISM: Staphylococcus aureus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(3198)
US-09-815-242-8519

Query Match 3.5%; Score 42.4; DB 10; Length 3198;
Best Local Similarity 53.7%; Pred. No. 4.1;
Matches 88; Conservative 0; Mismatches 76; Indels 0; Gaps 0;

```
Qy    381 TAAATGAATCTACGAATGGATGATAAAGTTCATGAATATTTTTGTTACTTCTGCAGTCAG 440
          ||| || |||    |||   |  | |  ||||| || | |||     ||      |||
Db   2608 TAAGTGCATCAGGCAATTCTTCTTTATCTTCATACATCTGTTTTGGCGTTAAAATATCAA 2549

Qy    441 ATCATGAGTTATTGAGTCTATTGTTTTTTTAAGCCTGTTTCAGATGATCCATCATCAGTA 500
          || |  |   ||   ||  | |||||   | | | |    ||  ||||| ||     | |
Db   2548 ATAAAAAACCATCTTGTTCAATGTTTAAATCACCATCCAACACTTGATCAATAGCTTGCA 2489

Qy    501 ACAACATACACGGTGTAGTCCCAAATCCATCATATGCACCTTCT 544
          |||||  |  |  ||||  | |||||  ||||| | || ||  ||
Db   2488 ACAACGTTGAACGTGTTTTACCAAAAGCATCAAACGCTCCCACT 2445
```

RESULT 8
US-10-239-676-127
; Sequence 127, Application US/10239676
; Publication No. US20030082609A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with Gene Regulation
; FILE REFERENCE: 5013.1003
; CURRENT APPLICATION NUMBER: US/10/239,676
; CURRENT FILING DATE: 2002-09-24
; PRIOR APPLICATION NUMBER: PCT/EP01/03968
; DE 10019058.8
; DE 10019173.8
; DE 10032529.7
; DE 10043826.1
; PRIOR FILING DATE: 2001-04-06
; 2000-04-06
; 2000-04-07
; 2000-06-30
; 2000-09-01
; NUMBER OF SEQ ID NOS: 228
; SEQ ID NO 127
; LENGTH: 6282
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-239-676-127

Query Match 3.3%; Score 40.4; DB 9; Length 6282;
Best Local Similarity 49.5%; Pred. No. 18;
Matches 104; Conservative 0; Mismatches 106; Indels 0; Gaps 0;

```
Qy    976 TGTAAGGTTGCATACAATACTTTTTTTTCAGACTTTACTTATGGTATTATACTGAATATG 1035
          |||  ||||  |   |||  |||| |||  |  |||      || |   ||  |   | ||
Db   2769 TGTTTGGTTTTAAGTAATTTTTTTGTTTTGGTTTTTTAGAGTGTTGGGATTATAGGTGTG 2828

Qy   1036 TTATTGCTGTTATAGTAGTTGAGTGACGTTTGAGGGAATTTCTAGTCCGTTAATCTTGTA 1095
               |    | ||||        |    ||| || ||| || || |    |  | |  ||
Db   2829 AGTTATTATATTTAGTTTAGTTTTAGATTTTTAGAGAAATTTTAATAGTATTTTATATTA 2888

Qy   1096 CTCAGTGTGTCTACTTTTCAAAAAAGTCAGTTTTTCAGTCTCTAAAACACATTTAAATAA 1155
               |  | | |  |||  ||||||  | |  |||| ||| | || |  | || | ||
Db   2889 TAGTGATTTTTTTTTTTAAAAAAAATTTATGTTTTTAGTATGTATAGTAAATATTAAAGT 2948

Qy   1156 GAGTTTCTTTGCCCATCTTTTGTTCCTCAT 1185
          || |||    |    | |||| ||  | ||
Db   2949 GATTTTTAAAGAATTTTTTTTTTTTATTAT 2978
```

RESULT 9
US-10-239-676-35
; Sequence 35, Application US/10239676
; Publication No. US20030082609A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with Gene Regulation
; FILE REFERENCE: 5013.1003
; CURRENT APPLICATION NUMBER: US/10/239,676
; CURRENT FILING DATE: 2002-09-24
; PRIOR APPLICATION NUMBER: PCT/EP01/03968
; DE 10019058.8
; DE 10019173.8
; DE 10032529.7

```
;    DE 10043826.1
;  PRIOR FILING DATE: 2001-04-06
;    2000-04-06
;    2000-04-07
;    2000-06-30
;    2000-09-01
;  NUMBER OF SEQ ID NOS: 228
; SEQ ID NO 35
;   LENGTH: 12405
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (7895)
US-10-239-676-35

  Query Match             3.3%;  Score 40.2;  DB 9;  Length 12405;
  Best Local Similarity   48.8%;  Pred. No. 28;
  Matches  137;  Conservative    0;  Mismatches  143;  Indels    1;  Gaps    1;

Qy      58 TTGGCTTACTGTTGTGCTCAAAGCAACTTCATCATCATACAGTATGGTTTTGATATGCTC 117
           ||||   |  |||||  | |  | |  ||      | ||   ||   |||| || |
Db    3191 TTGGAAAATGGTTGTATTAATTGAATTTTTTATGTTTTATTTTAGATTTTTAATTTTAAG 3250

Qy     118 TTCCATTATCACTGAGCCTTATGATTATGTTTTACGAGCTTATAATATCACTGATGGTGA 177
           ||   ||   |       |||  || | | ||      |  ||   ||| | | | ||| |
Db    3251 TTTTGTTTATAGATTTGATTAAAATAAAGGTTATTATGAATAGGTTATTAATAAAGGTTA 3310

Qy     178 TTCAGTATTGTGATTATGTCCTTCGTTGATTATTCTGTTTCATACAAGTCGTGTAATTTG 237
           |  ||| ||||| ||||     |    | | |  | |||||  |   | | ||  |  |||
Db    3311 TAAAGTTTTGTGTTTATGATGTGTTATAAATGGTGTGTTTAGTTTTATTAGTTAATGTTG 3370

Qy     238 CTGTTTGTGACAGTACGATAGATCGACTCAACCTTC-TGAGGTATTAGTTGAAGTTCATG 296
            ||| | |||         |||       |     ||  |    | ||| || || || |
Db    3371 TTGTATTTGATGTCGTTTTAGTGATTTTTTTTGTTAGTTTATTGTTATTTAAAATTAAGT 3430

Qy     297 TAAATTAGCTTTGTTTATCATAGTAGCATTTGATTATTGAT 337
           |||||   ||||| |  ||  |     || |||||| |
Db    3431 AAAATTTTTTTTGTATGAAATTTTTTTGATTAATTATTGTT 3471


RESULT 10
US-10-239-676-95
; Sequence 95, Application US/10239676
; Publication No. US20030082609A1
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with Gene Regulation
;  FILE REFERENCE: 5013.1003
;  CURRENT APPLICATION NUMBER: US/10/239,676
;  CURRENT FILING DATE:  2002-09-24
;  PRIOR APPLICATION NUMBER: PCT/EP01/03968
;    DE 10019058.8
;    DE 10019173.8
;    DE 10032529.7
;    DE 10043826.1
;  PRIOR FILING DATE: 2001-04-06
;    2000-04-06
;    2000-04-07
;    2000-06-30
;    2000-09-01
;  NUMBER OF SEQ ID NOS: 228
; SEQ ID NO 95
;   LENGTH: 15732
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-239-676-95

  Query Match             3.3%;  Score 40;  DB 9;  Length 15732;
  Best Local Similarity   47.2%;  Pred. No. 35;
  Matches  154;  Conservative    0;  Mismatches  170;  Indels    2;  Gaps    1;

Qy     194 TGTCCTTCGTTGATTATTCTGTTTCATACAAGTCGTGTAATTTGCTGTTTGTGACAGTAC 253
           |||  |   ||  || |  | ||| ||   | |  |  | |||  | ||| |     ||
Db   10184 TGTAATGTTTTTTTTTTAGTTTTGATTTTATTATTTGAGTTTTTTTTTTTTTTTTTTTAT 10243

Qy     254 GATAGATCGACTCAACCTTCTGAGGTATTAGTTGAAGTTCATGTAAATTAGCTTTGTTTA 313
           | ||  |    | ||  ||     |||||| ||  | ||   | ||||||  ||| |||
Db   10244 GTTAATTTAGTTAAAGTTTTGTTAGTATTATTTATATTTTTAGAAAATTAATTTTTTTTT 10303

Qy     314 TCATAGTAGCATTTGATTATTGATGCTCTGTAGCTAATGATAAGCCATTGGAGGGAAGCA 373
           | || |     ||| ||||||  |    | |   |  | |      ||  |
Db   10304 TTATTGATTTTTTTTATTATTTTTTTAATTTTAATTTTATTTTTAATTTTTATTTTTTTT 10363

Qy     374 AGCTTTCTAAATGAATCTACGAATGGATGATAAAGTTCATGAATATTTTTGTTACTTCTG 433
             ||| | |  |  | |       |  || |  | ||  | ||  |||||  |    | |
Db   10364 TTTTTTTTGATGGTTTTTGGTTTAGTTTGTTTTATTTTTTTAAGTTTTTTATGG--TGTA 10421

Qy     434 CAGTCAGATCATGAGTTATTGAGTCTATTGTTTTTTTAAGCCTGTTTCAGATGATCCATC 493
           |||  |||| ||   ||   |||| | || | |    |      |||||  |   |  |  ||
Db   10422 AAGTTAGATTATTGATTTGAGAGTTTTTTTTGTAAATGTAGTTGTTTAAATTTGTAAATT 10481

Qy     494 ATCAGTAACAACATACACGGTGTAGT 519
           ||   ||  |  |||     |||| |
Db   10482 ATTTTTATTATTATATTTTTTGTATT 10507


RESULT 11
US-09-960-352-8406
; Sequence 8406, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
;  APPLICANT: Warren, Wesley C.
;  APPLICANT:  Tao, Nengbing
;  APPLICANT:  Byatt, John C.
;  APPLICANT:  Mathialagan, Nagappan
;  TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION
;  TITLE OF INVENTION:  MUSCLE AND FAT DEPOSITION
;  FILE REFERENCE: 16511.006/37-21(10298)C
;  CURRENT APPLICATION NUMBER: US/09/960,352
;  CURRENT FILING DATE:  2001-09-24
;  NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 8406
;   LENGTH: 426
;   TYPE: DNA
;   ORGANISM: Bos taurus
;   OTHER INFORMATION: Clone ID: 36-LIB3058-032-Q1-K1-A12
US-09-960-352-8406

  Query Match             3.3%;  Score 39.6;  DB 10;  Length 426;
  Best Local Similarity   49.1%;  Pred. No. 6.9;
  Matches  105;  Conservative    0;  Mismatches  109;  Indels    0;  Gaps    0;

Qy     449 TTATTGAGTCTATTGTTTTTTTAAGCCTGTTTCAGATGATCCATCATCAGTAACAACATA 508
           || ||   | | || ||||||||   ||  ||   ||  |  ||  | | |    |  ||
Db     212 TTTTTTTTTTTTTTATTTTTTTATTTCTTATTTTTATTTTTTATTTTTATTTTTATTTA 271

Qy     509 CACGGTGTAGTCCCAAATCCATCATATGCACCTTCTTTTCTTCAATTTGGTCTTGTTTTT 568
             | |  |  | |    ||  |   | ||| || | || | |||  | || |||||
Db     272 TTATTTTTTTTATCTATATAATTTTTATTATCTTTTTATTTTTATTTTTTTTTTTTTTT 331

Qy     569 TTTTTTTCATGATGTCATTGAATTATTCAAGAAGTCACTTCGAGCATAATGATTTTTCAA 628
           | ||||| ||    | ||  | |||||  |    |   ||      |  |   ||||  |
Db     332 TATTTTTTATTTATTTATATATTTATTTTATTTTTTTTTTTTTTATTTTTTTTATTTTTTA 391

Qy     629 AATCCACCTTTGTTCAAGCACTACCACGTCTTTT 662
           |  |. |||||| |     ||   | | ||||
Db     392 TTTTTATTTTTGTTTATTTTTTATATCATTTTTT 425
```

```
RESULT 12
US-10-005-530-42/c
; Sequence 42, Application US/10005530
; Publication No. US20030026795A1
; GENERAL INFORMATION:
;  APPLICANT: Isaac, Barbara G.
;  APPLICANT:  Greenplate, John T.
;  APPLICANT:  Purcell, John P.
;  APPLICANT:  Romano, Charles P.
;  TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR CONTROLLING INSECTS
;  FILE REFERENCE: 11899.0022.DVUS01 (MOBT:022--2)
;  CURRENT APPLICATION NUMBER: US/10/005,530
;  CURRENT FILING DATE:  2001-10-26
;  PRIOR APPLICATION NUMBER: 09/063,733
;  PRIOR FILING DATE: 1998-04-21
;  PRIOR APPLICATION NUMBER: 60/044,504
;  PRIOR FILING DATE: 1997-04-21
;  NUMBER OF SEQ ID NOS: 58
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 42
;   LENGTH: 2093
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Synthetic Polynucleotide
US-10-005-530-42

  Query Match             3.3%;  Score 39.6;  DB 9;  Length 2093;
  Best Local Similarity   51.1%;  Pred. No. 16;
  Matches   93;  Conservative    0;  Mismatches   89;  Indels    0;  Gaps    0;

Qy      13 GTATTTATCTATAGCATTAGAAACTATAAGAGTTGTTAGCTTCACTTGGCTTACTGTTGT 72
           | ||| || ||   ||| |     | |     |  |||       |||    ||||| |
Db    2077 GGATTCATATACCTCATCAATTCTTTTTCACCTATTTATGCATCTTTGCAACACTGTAGA 2018

Qy      73 GCTCAAAGCAACTTCATCATCATACAGTATGGTTTTGATATGCTCTTCCATTATCACTGA 132
             |  | ||||||  |  ||   || |||||| |||   ||| |||  |    |||   |
Db    2017 CCAGATAGCAACACCTACAAGCTATAGTATGCTTTGCGTATTCTCAACAGCGATCTTGGT 1958

Qy     133 GCCTTATGATTATGTTTTACGAGCTTATAATATCACTGATGGTGATTCAGTATTGTGATT 192
            ||| | ||| ||| || |     |||  |||| |   || | |  |  ||||   ||
Db    1957 CACTTCTCATTGTGTGTTCCACAGATATCCTATCCCGCCTGCTTACCCCCTATTCACTTT 1898

Qy     193 AT 194
           |
Db    1897 CT 1896


RESULT 13
US-09-790-988-1/c
; Sequence 1, Application US/09790988
; Patent No. US20020127687A1
; GENERAL INFORMATION:
;  APPLICANT: SHIGENOBU, SHUJI
;  APPLICANT:  WATANABE, HIDEMI
;  APPLICANT:  HATTORI, MASAHIRA
;  APPLICANT:  SAKAKI, YOSHIYUKI
;  TITLE OF INVENTION: GENOME DNA OF BACTERIAL SYMBIONT OF APHIDS
;  FILE REFERENCE: 081356/0159
;  CURRENT APPLICATION NUMBER: US/09/790,988
;  CURRENT FILING DATE:  2001-02-23
;  PRIOR APPLICATION NUMBER: JP2000-107160
;  PRIOR FILING DATE: 2000-04-07
;  NUMBER OF SEQ ID NOS: 7
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
;   LENGTH: 640681
;   TYPE: DNA
;   ORGANISM: Buchnera sp.
US-09-790-988-1

  Query Match             3.3%;  Score 39.6;  DB 10;  Length 640681;
  Best Local Similarity   50.0%;  Pred. No. 2.8e+02;

  Matches   99;  Conservative    0;  Mismatches   99;  Indels    0;  Gaps    0;

Qy     102 TGGTTTTGATATGCTCTTCCATTATCACTGAGCCTTATGATTATGTTTTACGAGCTTATA 161
           | ||||| || || | ||  |  |  |  |  | || ||||   | |||   | |||
Db  202333 TTGTTTTTATCTGATATTGAACAACTAGAAAAACAAATAATTATCAATAACGCTTTAATA 202274

Qy     162 ATATCACTGATGGTGATTCAGTATTGTGATTATGTCCTTCGTTGATTATTCTGTTTCATA 221
             || | | ||  | || |  ||| |   ||           | |||| | | ||| ||
Db  202273 TAATTATTCATTATAATACCCTATCGCAGTTTAAATAAAAAATAATTAATATTTTTTATT 202214

Qy     222 CAAGTCGTGTAATTTGCTGTTTGTGACAGTACGATAGATCGACTCAACCTTCTGAGGTAT 281
           ||  || | | || | || | |   | | || | || ||  |    |  ||      | |
Db  202213 CATATCCTATGATGTTCTTTCTAAAAAAATATGTTATATAAAAAATATTTTAATTAATTT 202154

Qy     282 TAGTTGAAGTTCATGTAA 299
           |||||    |   | |||
Db  202153 TAGTTATCTTATTTTTAA 202136


RESULT 14
US-10-239-676-26
; Sequence 26, Application US/10239676
; Publication No. US20030082609A1
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with Gene Regulation
;  FILE REFERENCE: 5013.1003
;  CURRENT APPLICATION NUMBER: US/10/239,676
;  CURRENT FILING DATE:  2002-09-24
;  PRIOR APPLICATION NUMBER: PCT/EP01/03968
;    DE 10019058.8
;    DE 10019173.8
;    DE 10032529.7
;    DE 10043826.1
;  PRIOR FILING DATE: 2001-04-06
;    2000-04-06
;    2000-04-07
;    2000-06-30
;    2000-09-01
;  NUMBER OF SEQ ID NOS: 228
; SEQ ID NO 26
;   LENGTH: 9293
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-239-676-26

  Query Match             3.2%;  Score 39.2;  DB 9;  Length 9293;
  Best Local Similarity   45.7%;  Pred. No. 42;
  Matches  137;  Conservative    0;  Mismatches  163;  Indels    0;  Gaps    0;

Qy     435 AGTCAGATCATGAGTTATTGAGTCTATTGTTTTTTTAAGCCTGTTTCAGATGATCCATCA 494
           ||  || | | |   || |||||  |   | | ||  |  | ||   ||   |   |
Db    6721 AGGGAGTTAAGGTTGTAGTGAGTTAAGATTATGTTATTGTATTTTAGTGAGATTTTGTTT 6780

Qy     495 TCAGTAACAACATACACGGTGTAGTCCCAAATCCATCATATGCACCTTCTTTTCTTCAAT 554
           | | ||| || |   | |      |    || |  |  || |     || |  | || | |
Db    6781 TAATTAATAATAATAAAGTAAATGGGGGAAGTAAAGAATTTAGTTTTTTTGGTATTAAGT 6840

Qy     555 TTGGTCTTGTTTTTTTTTTTTCATGATGTCATTGAATTATTCAAGAAGTCACTTCGAGCA 614
           || || |  |||||| |||||    |||  | |   || | |         | ||| |
Db    6841 TTAGTATAATTTTTTATTTTTTTATATGATTTGGGGGTAATATAAGGAGAGAATAGAGAA 6900

Qy     615 TAATGATTTTTCAAAATCCACCTTTGTTCAAGCACTACCACGTCTTTTCATCTAGCCCAC 674
           | |  ||||| |||       || ||    |  |     | |||  |  |     |
Db    6901 AAGTATTTTTTAAAATATTTTGATTTTTGGTTTATAATTTTTTATTTAAAATTTTGGTAG 6960

Qy     675 AACCGTGGTGGAGGATCTAGAATTTTCATGAAAGGATTCAAAATTTACAAACATATATAT 734
             || |  || || || |||||| |  ||||  ||  | |     ||   || || |
Db    6961 TTAGATGTTTTAGAATTTAAAATTTTTAGAAAAGATTTTTAGAAAAGAAATTTTAAATTT 7020
```

```
RESULT 15
US-10-239-676-129
; Sequence 129, Application US/10239676
; Publication No. US20030082609A1
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with Gene Regulation
;  FILE REFERENCE: 5013.1003
;  CURRENT APPLICATION NUMBER: US/10/239,676
;  CURRENT FILING DATE:  2002-09-24
;  PRIOR APPLICATION NUMBER: PCT/EP01/03968
;    DE 10019058.8
;    DE 10019173.8
;    DE 10032529.7
;    DE 10043826.1
;  PRIOR FILING DATE: 2001-04-06
;    2000-04-06
;    2000-04-07
;    2000-06-30
;    2000-09-01
;  NUMBER OF SEQ ID NOS: 228
; SEQ ID NO 129
;   LENGTH: 6306
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-239-676-129

  Query Match             3.2%;  Score 39;  DB 9;  Length 6306;
  Best Local Similarity   50.3%;  Pred. No. 38;
  Matches   96;  Conservative    0;  Mismatches   95;  Indels    0;  Gaps    0;

Qy     974 TTTGTAAGGTTGCATACAATACTTTTTTTTCAGACTTTACTTATGGTATTATACTGAATA 1033
           ||| | |  ||   |    |  |||| |||  ||  | |  |    |  |||| || |||
Db    2679 TTTTTTATTTTTTTTTTTTTTTTTTTGTTTAGGATGTAATATGGAATTATATATTGTATA 2738

Qy    1034 TGTTATTGCTGTTATAGTAGTTGAGTGACGTTTGAGGGAATTTCTAGTCCGTTAATCTTG 1093
           |  |  |  ||||||  | |||   |   |||||    | | | |||   ||  | ||
Db    2739 TAGTTGTTATGTTATTTTGGTTTTTTTTAGTTTGGAATATTGTTTAGGTTTTTTTTATTT 2798

Qy    1094 TACTCAGTGTGTCTACTTTTCAAAAAAGTCAGTTTTTCAGTCTCTAAAACACATTTAAAT 1153
           |      | |   || |||| || |   |  |||  | | | | || ||    || || |
Db    2799 TTTATGATTTTGTTATTTTTGAAGAGTATAGGTTAGTTATTTTGTAGAATTATTTAAATT 2858

Qy    1154 AAGAGTTTCTT 1164
             || || | |
Db    2859 TGGATTTACGT 2869


Search completed: May 24, 2003, 09:17:35
Job time : 866 secs
```

GenCore version 5.1.5

OM nucleic - nucleic search, using sw model

Run on: May 24, 2003, 06:32:59 ; Search time 3210 Seconds
(without alignments)
11033.681 Million cell updates/sec

Title: US-09-701-926B-1
Perfect score: 1217
Sequence: 1 tttgaaatttatgtatttat.....gtcaacacaacacaacaaca 1217

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	471.4	38.7	2094	6	A50017	A50017 Sequence 1
2	59.4	4.9	1141	6	AX083744	AX083744 Sequence
3	53	4.4	8952	6	AX251200	AX251200 Sequence
c 4	51.6	4.2	1141	6	AX083744	AX083744 Sequence
5	51.2	4.2	61052	2	AC117074	AC117074 Dictyoste
6	51.2	4.2	61052	2	AC123513	AC123513 Dictyoste
c 7	51.2	4.2	268147	2	AC116966	AC116966 Dictyoste
8	51	4.2	6531	6	AX345543	AX345543 Sequence
c 9	51	4.2	125403	9	AC008929	AC008929 Homo sapi
10	51	4.2	148193	9	AC027347	AC027347 Homo sapi
11	51	4.2	220146	2	AC034179	AC034179 Homo sapi
c 12	51	4.2	223228	2	AC116365	AC116365 Homo sapi
c 13	49.8	4.1	74539	9	HS1012F16	AL080274 Human DNA
14	49.8	4.1	127811	2	AC008375	AC008375 Homo sapi
c 15	49.8	4.1	130540	2	AC079417	AC079417 Mus muscu
c 16	49.8	4.1	163056	2	AL359968	AL359968 Homo sapi
17	49.8	4.1	173669	2	AC017110	AC017110 Homo sapi
c 18	49.8	4.1	179203	9	AC009194	AC009194 Homo sapi
19	49.8	4.1	189486	9	AC008733	AC008733 Homo sapi
c 20	49.8	4.1	201981	2	AC073640	AC073640 Homo sapi
21	49.4	4.1	196216	8	AF165818	AF165818 Guillardi
22	49.2	4.0	5647	6	AX252020	AX252020 Sequence
23	49.2	4.0	5647	6	AX346468	AX346468 Sequence
24	49.2	4.0	5647	6	AX348787	AX348787 Sequence
25	49.2	4.0	154028	9	HSAC002087	AC002087 Human BAC
c 26	48.8	4.0	107739	2	AC116979	AC116979 Dictyoste
27	48.4	4.0	15674	6	AX281366	AX281366 Sequence
28	48.4	4.0	15674	6	AX345265	AX345265 Sequence
29	48.4	4.0	15674	6	AX348946	AX348946 Sequence
c 30	48.2	4.0	146383	2	AC116367	AC116367 Oryza sat
c 31	48	3.9	894	11	CNS06KR0	AL403234 T3 end of
32	47.2	3.9	66441	3	PFMAL1P4	AL031747 Plasmodiu
33	47.2	3.9	253305	3	PFMAL3P7	AL034559 Plasmodiu
c 34	47	3.9	124820	2	AC117073	AC117073 Dictyoste
c 35	46.8	3.8	56152	2	AC116963	AC116963 Dictyoste
c 36	46.8	3.8	153837	2	AL772144	AL772144 Danio rer
c 37	46.6	3.8	107289	2	AC116923	AC116923 Dictyoste
38	46.6	3.8	140211	2	AC122656	AC122656 Rattus no
39	46.4	3.8	7218	6	I66494	I66494 Sequence 14
c 40	46.4	3.8	147666	2	AP004635	AP004635 Oryza sat
41	46.4	3.8	182200	2	AP005504	AP005504 Oryza sat
42	46.4	3.8	198572	10	AL732460	AL732460 Mouse DNA
43	46.2	3.8	863	11	CNS06EVQ	AL395628 T7 end of
44	46.2	3.8	19236	6	AX344818	AX344818 Sequence
45	46.2	3.8	349980	6	AX344560	AX344560 Sequence

ALIGNMENTS

RESULT 1
A50017
LOCUS A50017 2094 bp DNA linear PAT 07-MAR-1997
DEFINITION Sequence 1 from Patent WO9612813.
ACCESSION A50017
VERSION A50017.1 GI:2303200
KEYWORDS .
SOURCE unidentified.
ORGANISM unidentified
unclassified.
REFERENCE 1 (bases 1 to 2094)
AUTHORS Pedersen,H.F., Kreiberg,J.D. and Lund,M.
TITLE PROMOTER FROM A PLANT ALPHA-AMYLASE GENE
JOURNAL Patent: WO 9612813-A 1 02-MAY-1996;
DANISCO (DK)

```
COMMENT     Other publication AU 3075095 960515.
FEATURES             Location/Qualifiers
     source          1. .2094
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      618 a    357 c    369 g    750 t
ORIGIN

  Query Match           38.7%;  Score 471.4;  DB 6;  Length 2094;
  Best Local Similarity   77.5%;  Pred. No. 6.5e-86;
  Matches  680;  Conservative    0;  Mismatches  171;  Indels   26;  Gaps    8;

Qy       1 TTTGAAATTTATGTATTTATCTATAGCATTAGAAACTATAAGAGTTGTTAGCTTCACTTG 60
          ||||||||||||||| ||||| ||||||||||||||||||||||||||||||||||||||
Db     981 TTTGAAATTTATGTATATATCTGTAGCATTAGAAACTATAAGAGTTGTTAGCTTCACTTG 1040

Qy      61 GCTTACTGTTGTGCTCAAAGCAACTTCATCATCATACAGTATGGTTTTGATATGCTCTTC 120
           |||| ||||||||||||||||||   ||||||||||||||||||| ||||||||||||
Db    1041 TCTTATTGTTGTGCTCAAAGCAACT---TCATCATACAGTATGGTTTTTATATGCTCTTC 1097

Qy     121 CATTATCACTGAGCCTTATGATTATGTTTTACGAGCTTATAATATCACTGATGGTGATTC 180
           ||||||||| || |||||||||||||| | |||||||||||||||| ||||||||||||||
Db    1098 CATTATCACCGAACCTTATGATTATG-TGTACGAGCTTATAATATTACTGATGGTGATTC 1156

Qy     181 AGTATTGTGATTATGTCCTTCGTTGATTATTCTGTTTCATACAAGTCGTGTAATTTGCTG 240
           |||||| |||||||||||| | || ||||||||||||||||||||||||||||||||||||
Db    1157 AGTATTATGATTATGTCCTCCATTAATTATTCTGTTTCATACAAGTCGTGTAATTTGCTG 1216

Qy     241 TTTGTGACAGTACGATAGATCGACTCAACCTTCTGAGGTATTAGTTGAAGTTCATGTAAA 300
           |||||||  |||||||| || || |||||||||| ||| || |||||||||| |||||
Db    1217 TTTGTGATTGTACGATAAATTGATTCAACCTTCTGCGGTGTTGGTTGAAGTTCAAGTAAA 1276

Qy     301 TTAGCTTTGTTTATCATAGTAGCATTTGATTATTGATGCTCTGTAGCTAATGATAAGCCA 360
           |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1277 TTAGCTTTATTTATCATAGTAGCATTTGATTATTGATGCTCTGTAGCTAATGATAAGCCA 1336

Qy     361 TTGGAGGGAAGC-----------AAGCTTTCTAAATGAATCTACGAATGGATGATAAAGT 409
           ||| ||||||||           |   ||   ||||||||||||||||||||||||||||
Db    1337 TTGAAGGGAAGCAGAAATGGTAAAGCTTTCTAAAATGAATCTACGAATGGATGATAAAGT 1396

Qy     410 TCATGAATATTTTTGTTACTTCTGCAGTCAGATCATGAGTTATTGAGTCTATTGTTTTTT 469
           | |||||||| ||| |||||||| |||||| |||||||| |||||||| || |||||
Db    1397 TAATGAATATTGTTGATACTTCTGCAATCAGATTATGAGTTACTGAGTCTACTG-TTTTT 1455

Qy     470 TAAGCCTGTTTCAGATGATCCATCATCAGTAACAACATACACGGTGTAGTCCCAAATCCA 529
           |||||||||||||||||| |||||||  |||||||||  | |||||||    | |
Db    1456 TAAGCCTGTTTCAGATGATCGATCATCAACAACAACATATTCAGTGTAGT----AGACAT 1511

Qy     530 TCATATGCACCTTCTTTTCTTCAATTTGGTCTTGTTTTTTTTTTTTCATGATGTCATTGA 589
              ||| |||| | ||| ||| |     | |||| |       || |    ||
Db    1512 GATCGATCACTTTCTAATTTTCGATTATGCACCCTCTTTTCTCCAATTTGGTCGTCTTCT 1571

Qy     590 ATTATTCAAGAAGTCACTTCGAGCATAATGATTTTTCAAAATCCACCTTTGTTCAAGCAC 649
           || |||| || ||||||    | || ||  | |        ||||| ||    |||
Db    1572 TTTTTTCATGATGTCACT----GAATTATTCTCTGGTCGTCCCCACCATTCAGGAAGTCA 1627

Qy     650 TACCACGTCTTTTCATCTAGCCCACAACCGTGGTGGAGGATCTAGAATTTTCATGAAAGG 709
             |  |  |  |   | |  ||    |     |     |||||||||| ||| |
Db    1628 CTTCGAGCATAATGTGAAAACATCCACATTTTTCAAATCCAGCAGAATTTTCATCAAACG 1687

Qy     710 ATTCAAAATTTACAAACATATATATACACTATACACTATGAATCCACTAATACTAGATGG 769
           ||  |   |  | || ||||||||  |  | ||||||||||||| ||||||||
Db    1688 GGGTTCAACAT-TTACTACATGTATACACTCTGAAGTCTGAATCCACTAATTCTAGATGG 1746

Qy     770 TGCACCTGTGCCCCCACTCATGTGAAAGCCTATTCTCAATTTTTTATTTTCC-ACAACTT 828
           |||| |||||||||||| | |||||||| |||||||||||||||||||||| |||||||
Db    1747 TGCATCTGTGCCCCCACACTTGTGAAAGCTTATTCTCAATTTTTTATTTTCCAACAACTT 1806

Qy     829 AAATACAGACCGCACAACTCCCGTGTCTTGTGTGCTC 865
           ||| |||||| ||||||||||||||||||  | ||
Db    1807 GAATTCAGACCACACAACTCCCGTGTCTTGTACGGTC 1843
```

```
RESULT 2
AX083744
LOCUS       AX083744               1141 bp    DNA     linear   PAT 28-FEB-2001
DEFINITION  Sequence 22 from Patent WO0111061.
ACCESSION   AX083744
VERSION     AX083744.1  GI:13185472
KEYWORDS    .
SOURCE      synthetic construct.
  ORGANISM  synthetic construct
            artificial sequences.
REFERENCE   1  (bases 1 to 1141)
  AUTHORS   Kunst,L. and Clemens,S.
  TITLE     Regulation of embryonic transcription in plants
  JOURNAL   Patent: WO 0111061-A 22 15-FEB-2001;
            UNIVERSITY OF BRITISH COLUMBIA (CA)
FEATURES             Location/Qualifiers
     source          1. .1141
                     /organism="synthetic construct"
                     /db_xref="taxon:32630"
     promoter        1. .1141
                     /note="consensus sequence of A.t., L.a., and B.n. FAE1
                     promoters"
BASE COUNT      123 a     32 c     42 g    112 t    832 others
ORIGIN

  Query Match           4.9%;  Score 59.4;  DB 6;  Length 1141;
  Best Local Similarity   11.1%;  Pred. No. 0.034;
  Matches  120;  Conservative  398;  Mismatches  551;  Indels    9;  Gaps    3;

Qy      88 ATCATCATACAGTATGGTTTTGATATGCTCTTCCATTATCACTGAGCCTTATGATTATGT 147
           :| | :   ::::   :  ::: : :: :::   |  :: ::: ::||:::|   ::
Db      59 WTWARMYCKYRRWYNNKSRWWKGWYKKKWYBCANNTSBRYHARRWKDMKTAYBMTMTNKW 118

Qy     148 TTTACGAGCTTATAATATCACTGATGGTGATTCAGTATTGTGATTATGTCCTTCGTTGAT 207
           :|  ::::: : |:::|:::: :|    |:| :  :::::::| :: :       ::::
Db     119 GKTGWRHRYWRWRAMBDTVDHHYVTAMNNAWTTMCMMDKDDKRTRWWWKKNNNATGWDDD 178

Qy     208 TATTCTGTTTCATACAAGTCGTGTAATTTGC-TGTTTGTGACAGTACGATAGATCGACTC 266
           | ::::    | |: ::: ::|: :: :::: :| :: : :: :  : ::: :::   ::
Db     179 TKYHMWNNNGCBTVTWMVRYKTDRDWSBKRMNYGMBWWKNWSYDVTYYWWVWDDMCKRKV 238

Qy     267 AACCTTCTGAGGTATTAGTTGAAGTTCATGTAAATTAGCTTTGTTTATCATAGTAGCATT 326
           :: : |  ::: ::  ::|  ::::   : :|:|: :          |: :  :     :
Db     239 RRWVRTRGRMRNYMVAWBTAHRRRYNNGWTBAMAYRRWTMNNNNNNAKAMCKRAKYWGWN 298

Qy     327 TGATTATTGATGCTCTGTAGCTAATGATAAGCCATTGGAGGGAAGCAAGCTTTCTAAATG 386
           |   :::: |: : : :|    | :   ::: :: :||  :         : ::
Db     299 RABVNSTCTTWKSKTTKVRTSCWANNCRAGDANKDHKWWKWSAAMGVYWNNNNNNNWTYK 358

Qy     387 AATCTACGAATGGATGATAAAGTTCATGAATATTTTTGTTACTTCTGCAGTCAGATCATG 446
           | ::|  ::        :::|   :  :  : ::: :         | ::: :: ::   :
Db     359 KARHBARWDWVWHSAWKKWHANAAHYSRKKWTBYKRKTMVNNNNGTTMWKRMWAWYWKMD 418

Qy     447 AGTTATTGAGTCTATTGTTTTTTTAAGCCTGTTTCAGATGATCCATCATCAGTAACAACA 506
           :::: |:         | : :|: : :  :::::     ::|: : || :   |:: :
Db     419 MDWBGTYNNNNNGGRTYYGWTKNKKMWTYYKWKANNCKWRAWDHKTCTHNNTTWWKMKTY 478

Qy     507 TACACGGTGTAGTCCCAAATCCATCATATGCACCTTCTTTTCTTCAATTTGGTCTTGTTT 566
           :   : :  | |:::  || :: :::::::  :     ::::: | |::: ::  :::
Db     479 WNNCYWKSMTNGKSHRBAAAVYTWYMWWWRRYAHANNNNWDYWWKACTWYKYBVCSKWWN 538

Qy     567 TTTTTTTTTCATGATGTCATTGAATTATTCAAGAAGTCA------CTTCGAGCATAATGA 620
           :  ::|: ::    ::::::    ::: :: : :|         :: :::  ::: :::
Db     539 NYAAWYTKSSWNYTSRYYRWKTNNSWRWRSDTRSMGRANNYARABHYGYKWNTRWWBWSH 598

Qy     621 TTTTTCAAAATCCACCTTTGTTCAAGCACTACCACGTCTTTTCATCTAGCCCACAACCGT 680
           |:::: | ||:: ::: :: : : :: :::|    |
Db     599 TWBHBRAGAAHYWMBMMYBAKCHCMKAWYKAKKYAGAGGSNNNNNNNNNNNNNNNNATCA 658

Qy     681 GGTGGAGGATCTAGAATTTTCATGAAAGGATTCAAAATTTACAAACATATATATACACTA 740
           :::  | :::: :  |:::::  : |     |:   ::  :  | : :   ::
Db     659 RDDYYAASRWYAMANAKWYYYKBAANNAYYTHANNWWGCWNNATDTRRTMWKNNNNNNAG 718
```

```
Qy    741 TACACTATGAATCCACTAATACTAGATGGTGCACCTGTGCCCCCACTCATGTGAAAGCCT 800
Db    719 TWKNNNNNNAKNASAAKNYAAAAVKAAKKHWRWANKWAMRGWHADAAABTTDKRNNGAYT 778

Qy    801 ATTCTCAATTTTTTATTTTCCACAACTTAAATACAGACCGCACAACTCCCGTGTCTTGTG 860
Db    779 KYTTTNNNNTYRGVVTNTAARDGWANNNNNNNNNNNNNNNGWSDMWVTWWAYANYGTNNNN 838

Qy    861 TGCTCGTCGCTCAGCATGCAAGTCGAGAAAAGAAAGACCAAAACAATGAAAACTTTACGA 920
Db    839 NNNNNNAYAWWTNKWYYTTDDRWRBAYTNNNNNNRMAYY--GAYADDYAYYMSDTCDAWM 896

Qy    921 AAAATCAAAAAGTTGAAGGACTTTAACGTCGAGATCTCTCGTAGAAAACCTCTTTTGTAA 980
Db    897 KWDATKMNNATTYNRGTAWRTNNNNNNMTMKTKYYBHAAWNNNNNNGKMCTAHTWWVCKA 956

Qy    981 GGTTGCATACAATACTTTTTTTTCAGACTTTACTTATGGTATTATACTGAATATGTTATT 1040
Db    957 TKTTKGCWMNCTTTCRKYKNNCTWYTWMTTTRTTWYAATRWKTNNATGSMTRCNATGWKN 1016

Qy   1041 GCTGTTATAGTAGTTGAGTGACGTTTGAGGGAATTTCTAGTCCGTTAATCTTGTACTCAG 1100
Db   1017 NNYWTGWKTRWTAYRMATRWMKAWWKVMATGSWNTNSYARWAYKTRAYKGWYYNACAWRW 1076

Qy   1101 TGTGTCTACTTTTCAAAAAAGTCAGTTTTTCAGTCTCTAAAACACATTTAAATAAGAG 1158
Db   1077 RWGKATCYMTDNAWWTACATSWMATHKYNWHMCKCNNNNNNNNTMMRAMAMAAANCDG 1134
```

RESULT 3
AX251200

LOCUS	AX251200 8952 bp DNA linear PAT 05-OCT-2001
DEFINITION	Sequence 168 from Patent WO0168912.
ACCESSION	AX251200
VERSION	AX251200.1 GI:15984623
KEYWORDS	.
SOURCE	synthetic construct.
ORGANISM	synthetic construct artificial sequences.
REFERENCE	1 (bases 1 to 8952)
AUTHORS	Olek,A., Piepenbrock,C. and Berlin,K.
TITLE	Diagnosis of diseases associated with tumor suppressor genes and oncogenes
JOURNAL	Patent: WO 0168912-A 168 20-SEP-2001; Epigenomics AG (DE)

FEATURES	Location/Qualifiers
source	1. .8952 /organism="synthetic construct" /db_xref="taxon:32630" /note="chemically treated genomic DNA (Homo sapiens)"

BASE COUNT 2404 a 162 c 1905 g 4481 t
ORIGIN

Query Match 4.4%; Score 53; DB 6; Length 8952;
Best Local Similarity 50.2%; Pred. No. 0.46;
Matches 157; Conservative 0; Mismatches 155; Indels 1; Gaps 1;

```
Qy     40 AAGAGTTGTTAGCTTCACTTGGCTTACTGTTGTGCTCAAAGCAACTTCATCATCATACAG 99
Db   6392 AAGATTTTTTATTTTTTTTTGGGTTTGTTTTGTGTTTTTAGGAATTTGTTTATGTTATTT 6451

Qy    100 TATGGTTTTGATATGCTCTTCCATTATCACTGAGCCTTATGATTATGTTTTACGAGCTTA 159
Db   6452 AGAGTATTTAATTTGTTGGTATTTAATTGTTTA-TATTATTTTTATATAATATTTTTTAT 6510

Qy    160 TAATATCACTGATGGTGATTCAGTATTGTGATTATGTCCTTCGTTGATTATTCTGTTTCA 219
Db   6511 TTTTATATAGGTTGGTGGTAATGTGTTTTTTTTTATTTTGATTTTAGTTATTTTATTTTT 6570

Qy    220 TACAAGTCGTGTAATTTGCTGTTTGTGACAGTACGATAGATCGACTCAACCTTCTGAGGT 279
Db   6571 TTGTTTTTTTGTTGTTTGTTTGTTTGTTTTATTGGTTAGTTTAGTTAAAGGTTTTTAAGT 6630
```

```
Qy    280 ATTAGTTGAAGTTCATGTAAATTAGCTTTGTTTATCATAGTAGCATTTGATTATTGATGC 339
Db   6631 TTTGGTGATAAATATTAAAATTTGGTTTTATTTATTTTTTTTGGTTTTTTTTTTTTTTTT 6690

Qy    340 TCTGTAGCTAATG 352
Db   6691 TTTAGATTTATTG 6703
```

RESULT 4
AX083744/c

LOCUS	AX083744 1141 bp DNA linear PAT 28-FEB-2001
DEFINITION	Sequence 22 from Patent WO0111061.
ACCESSION	AX083744
VERSION	AX083744.1 GI:13185472
KEYWORDS	.
SOURCE	synthetic construct.
ORGANISM	synthetic construct artificial sequences.
REFERENCE	1 (bases 1 to 1141)
AUTHORS	Kunst,L. and Clemens,S.
TITLE	Regulation of embryonic transcription in plants
JOURNAL	Patent: WO 0111061-A 22 15-FEB-2001; UNIVERSITY OF BRITISH COLUMBIA (CA)

FEATURES	Location/Qualifiers
source	1. .1141 /organism="synthetic construct" /db_xref="taxon:32630"
promoter	1. .1141 /note="consensus sequence of A.t., L.a., and B.n. FAE1 promoters"

BASE COUNT 123 a 32 c 42 g 112 t 832 others
ORIGIN

Query Match 4.2%; Score 51.6; DB 6; Length 1141;
Best Local Similarity 11.6%; Pred. No. 1.3;
Matches 117; Conservative 353; Mismatches 531; Indels 4; Gaps 2;

```
Qy     90 CATCATACAGTATGGTTTTGATATGCTCTTCCATTATCACTGAGCCTTATGATTATGTTT 149
Db   1060 MRTWYTRSNANWSCATKBMWWTMKWYATKYRTAWYAMWCAWRNNNMWCATNGYAKSCATN 1001

Qy    150 TACGAGCTTATAATATCACTGATGGTGATTCAGTATTGTGATTATGTCCTTCGTTGATTA 209
Db   1000 NAMWYATTRWAAYAAAKWARWAGNNMRMYGAAAGNKWGCMAAMATMGBWWADTAGKMCNN 941

Qy    210 TTCTGTTTCATACAAGTCGTGTAATTTGCTGTTTGTGACAGTACGATAGATCGACTCAAC 269
Db    940 NNNNWTTDVRRMAMKAKNNNNNNAYWTACYNRAATNNKMATHWMKWTHGAHSKRRTRHHT 881

Qy    270 CTTCTGAGGTATTAGTTGAAGTTCATGTAAATTAGCTTTGTTTATCATAGTAGCATTTGA 329
Db    880 RTCRRTKYNNNNNNARTVYWYHHAARRWMNAWWTRTNNNNNNNNNNACRNTRTWWABWKH 821

Qy    330 TTATTGATGCTCTGTAGCTAATGATAAGCCATTGGAGGGAAGCAAGCTTTCTAAATGAAT 389
Db    820 SWCNNNNNNNNNNNNNNTWCHYTTANABBCYRANNNNAAARMARTCNNYMHAAVTTTHTD 761

Qy    390 CTACGAATGGATGATAAAGTTCATGAATATTTTTGTTACTTCTGCAGTCAGATCATGAGT 449
Db    760 WCYKTWMNTWYWDMMTTMBTTTTRNMTTSTNMTNNNNNNMWACTNNNNNNMWKAYYAHAT 701

Qy    450 TATTGAGTCTATTGTTTTTTTAAGCCTGTTTCAGATGATCCATCATCAGTAACAACATAC 509
Db    700 NNWGCWWNNTDARRTNNTTVMRRRWMTNTKTRWYSTTRRHHYTGATNNNNNNNNNNNNNN 641

Qy    510 ACGGTGTAGTCCCAAATCCATCATATGCACCTTCTTTTCTTCAATTTGGTCTTGTTTTTT 569
Db    640 NNSCCTCTRMMTMRWTMKGDGMTVRKKVKWRDTTCTYVDVWADSWVWWYANWMRCRDVTY 581

Qy    570 TTTTTTCATGATG-TCATTGAATTATTCAAGAAGTCACTTCGAGCATAATGATTTTTCAA 628
Db    580 TRNNTYCKSYAHSYWYWSNNAMWYRRYSARNWSSMARWTTRNNWWMSGBVRMRWAGTMWW 521
```

```
Qy     629 AATCCACCTTTGTTCAAGCACTACCACGTCTTTTCATCTAGCCCACAACCGTGGTGGAGG 688
           :::     :|  :: : : | :   ::: :   ::::       : ::     : | ::
Db     520 RHWNNNNTDTRYYWWWKRWARBTTTVYDSMCNAKSMWRGNNWRAMKMWWAANNDAGAMDH 461

Qy     689 ATCTAGAATTTTCATGAAAGGATTCAAAATTTACAAACATATATATACACTATACACTAT 748
           :|:::|  | :  :   ::  |:  :|        :||::::::::   :  : |
Db     460 WTYWMGNNTMWMRRAWKMMNMAWCRRAYCCNNNNNRACVWHKHKMWRWTWKYMWKAACNN 401

Qy     749 GAATCCACTAATACTAGATGGTGCACCTGTGCCCCCACTCATGTGAAAGCCTATTCTCAA 808
           : : : :|:::: ::|  |: :: |:: : : : :::     :       : :
Db     400 NNBKAMYMRVAWMMYSRDTTNTDWMMWTSDWBWHWYTVDYTMMRAWNNNNNNNWRBCKTT 341

Qy     809 TTTTTTATTTTCCACAACTTAAATACAGACCGCACAACTCCCGTGTCTTGTGTGCTCGTC 868
           : :: ::  |:| :   |:  |: :| :::  | |  :: : : :  |  |: :|
Db     340 SWMWWMDHMNTHCTYGNNTWGSAYBMAAMSMWAAGASNBVTYNWCWRMTYMGKTMTNNNN 281

Qy     869 GCTCAGCATGCAAGTCGA---GAAAAGAAAGACCAAAACAATGAAAACTTTACGAAAAAT 925
           : : ::|: :|    :   : |::  :   :  | :: ::: :  :::   ::| :
Db     280 NNKAWYYRTKTVAWCNNRYYYDTAVWTBKRNYKYCYAYBWYYBMYMGKHHWBWWRRABHR 221

Qy     926 CAAAAAGTTGAAGGACTTTAACGTCGAGATCTCTCGTAGAAAACCTCTTTTGTAAGGTTG 985
           :: ::::: :    ::::::|:::::   : |   : :::|::::| |    ::  :
Db     220 SWNMWWVKCRNKYMVSWHYHAMRYBKWABAVGCNNNWKDRMAHHHWCATNNNMMWWWYAY 161

Qy     986 CATACAATACTTTTTTTTCAGACTTTACTTATGGTATTATACTGAATATGTTATTGCTGT 1045
           :::::      :|  :| :::::: :: |::: :::::  :|:  |:|:: || : :: :
Db     160 MHHMHKKGKAAWTNNKTABRDDHBAHVKTYWYWRYDYWCAMCWMNAKAKVRTAMKHMWYY 101

Qy    1046 TATAGTAGTTGAGTGACGTTTGAGGGAATTTCTAGTCCGTTAATC 1090
           |:  : |- ||::: ::::   : :    ::: :|  : : ::::
Db     100 TDRYVSANNTGVRWMMMRWCMWWYSMNNRWYYRMGRKYTWAWMYS 56
```

RESULT 5
AC117074
LOCUS AC117074 61052 bp DNA linear HTG 06-APR-2002
DEFINITION Dictyostelium discoideum chromosome 2 map 2779865-2840915 strain AX4, *** SEQUENCING IN PROGRESS ***, in ordered pieces.
ACCESSION AC117074
VERSION AC117074.1 GI:20066241
KEYWORDS HTG; HTGS_PHASE2.
SOURCE Dictyostelium discoideum.
ORGANISM Dictyostelium discoideum
Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE 1 (bases 1 to 61052)
AUTHORS Gloeckner,G., Eichinger,L., Szafranski,K., Pachebat,J., Dear,P., Lehmann,R., Baumgart,C., Parra,G., April,J.F., Guigo,R., Kumpf,K., Tunggal,B., Cox,E., Quail,M.A., Platzer,M., Rosenthal,A. and Noegel,A.A.
TITLE Sequence and Analysis of Chromosome 2 of Dictyostelium
JOURNAL Unpublished
REMARK The Dictyostelium Genome Sequencing Consortium
REFERENCE 2 (bases 1 to 61052)
AUTHORS Baumgart,C.
TITLE Direct Submission
JOURNAL Submitted (06-APR-2002) Genome Analysis, Institute of Molecular Biotechnology, Beutenberstr. 11, Jena 07745, Germany
COMMENT CDS predictions from GeneID may contain errors. Further Information is available from IMB Jena, Department of Genome Analysis (http://genome.imb-jena.de/dictyostelium/)
and the Univerity Colonge, Institute for Biochemistry I (http://www.uni-koeln.de/dictyostelium/project.shtml
Funding
Agency : Deutsche Forschungsgemeinschaft (DFG).
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
FEATURES Location/Qualifiers
source 1. .61052
/organism="Dictyostelium discoideum"
/strain="AX4"
/db_xref="taxon:44689"
/chromosome="2"
/map="2779865-2840915"
BASE COUNT 25445 a 6805 c 6328 g 22474 t
ORIGIN

```
Query Match              4.2%;  Score 51.2;  DB 2;  Length 61052;
Best Local Similarity   48.0%;  Pred. No. 0.76;
Matches  146;  Conservative    0;  Mismatches  158;  Indels    0;  Gaps    0;

Qy     886 AGAAAAGAAAGACCAAAACAATGAAAACTTTACGAAAAATCAAAAAGTTGAAGGACTTTA 945
           | |||| ||| |  |||| ||||||||     |  ||  |  |||||    ||  |
Db   28474 AAAAAAAAAAAAAAAAAAAAATGAAAAAAAAAAAAAATGAAAAAAAAAAAAAATTAAAAAT 28533

Qy     946 ACGTCGAGATCTCTCGTAGAAAACCTCTTTTGTAAGGTTGCATACAATACTTTTTTTTCA 1005
           |  |    |  | |     | ||  | |||   ||   |  |  |||||||  ||||||
Db   28534 AAATAATAAATTTTGACCCACAAATTATTTGTAAATAATTAAACCAATACTGTTTTTTTT 28593

Qy    1006 GACTTTACTTATGGTATTATACTGAATATGTTATTGCTGTTATAGTAGTTGAGTGACGTT 1065
            |||  | ||   | |    |  ||| | |||  | || |  || ||   || | |
Db   28594 TTTTTTTTTAATAAAAGTTATATTTTTATTTAATTTTTTTTTTTTTTATTTCTTTGTCCTC 28653

Qy    1066 TGAGGGAATTTCTAGTCCGTTAATCTTGTACTCAGTGTGTCTACTTTTCAAAAAAGTCAG 1125
            | |  | |  |   ||  | || | |  |  | | | |  |||  ||||||   |
Db   28654 CTCTTTGAATATTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTAAAAAAAAAAAAA 28713

Qy    1126 TTTTTCAGTCTCTAAAACACATTTAAATAAGAGTTTCTTTGCCCATCTTTTGTTCCTCAT 1185
            ||||  | | |  | | |  | || || |   ||||    |  |||| ||  |
Db   28714 ATTTTTTTTGTTTTTTAAAAAAATGAAAAAAATCAACTTTAAAAAATTTTTTTTTTTTTA 28773

Qy    1186 CCTA 1189
           | ||
Db   28774 CATA 28777
```

RESULT 6
AC123513
LOCUS AC123513 61052 bp DNA linear HTG 29-MAY-2002
DEFINITION Dictyostelium discoideum chromosome 2 map 2779865-2840915 strain AX4, *** SEQUENCING IN PROGRESS ***, in ordered pieces.
ACCESSION AC123513
VERSION AC123513.1 GI:21240650
KEYWORDS HTG; HTGS_PHASE2.
SOURCE Dictyostelium discoideum.
ORGANISM Dictyostelium discoideum
Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE 1 (bases 1 to 61052)
AUTHORS Gloeckner,G., Eichinger,L., Szafranski,K., Pachebat,J., Dear,P., Lehmann,R., Baumgart,C., Parra,G., April,J.F., Guigo,R., Kumpf,K., Tunggal,B., Cox,E., Quail,M.A., Platzer,M., Rosenthal,A. and Noegel,A.A.
TITLE Sequence and Analysis of Chromosome 2 of Dictyostelium
JOURNAL Unpublished
REMARK The Dictyostelium Genome Sequencing Consortium
REFERENCE 2 (bases 1 to 61052)
AUTHORS Baumgart,C.
TITLE Direct Submission
JOURNAL Submitted (29-MAY-2002) Genome Analysis, Institute of Molecular Biotechnology, Beutenberstr. 11, Jena 07745, Germany
COMMENT CDS predictions from GeneID may contain errors. Further Information is available from IMB Jena, Department of Genome Analysis (http://genome.imb-jena.de/dictyostelium/)
and the Univerity Colonge, Institute for Biochemistry I (http://www.uni-koeln.de/dictyostelium/project.shtml
Funding
Agency : Deutsche Forschungsgemeinschaft (DFG).
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
FEATURES Location/Qualifiers
source 1. .61052
/organism="Dictyostelium discoideum"

/strain="AX4"
/db_xref="taxon:44689"
/chromosome="2"
/map="2779865-2840915"

CDS join(448. .628,705. .803,875. .1134)
/note="ORF_ID:dd_02708"
/pseudo

CDS /codon_start=1
complement(join(1193. .1583,1653. .6117,6198. .6330))
/note="ORF_ID:dd_02705"
/codon_start=1
/product="ABC1 TRANSPORTER"
/protein_id="AAM44362.1"
/db_xref="GI:21240651"
/translation="MGEKSSQLKTLLKKNLLLKSKSKCGICCEIVFPIIIVLLVFAIL
VLVQLFKPNYDLIKTTQFSNRINENNIIIYGGKAGSLNVEQKGVIDMMKFQLSNELNK
SISEVDVYFKEINDKSEMENYFQINSTQVFGGIWFESNQLSSVANTTNTNTTTNSFKY
SIRLDSNSVLDSNKVKDNGIDSSDYLTKNWAYIQIAMDQAIFGYFGLDYKLVINGQRY
PDPYVELWQKWINGRESVFKSAGSVFVSAALLIFTFRLVTELVVEKETKIREGMSIMG
LNQYCYFISWIITSLVTALPIDLIIIVILKGSQVIHSTSWIIVIVTLILYLLTLQLLA
FIFSMFFDKSKFAGLLTFLTILLINICGIFIGEYEIINIHIKLLLCCIFSPIGIACSF
YIMSIRDIPDTISTINLNQVISEKQIIGTFVFNIIFYTFLIWYLDKIVKTEYGTKEPW
YFLFTKRYWTVGAGKINKKNKNKEYNYNDIESTIQNNNNDNIEMVPIEVRNKTTISIR
NLRKEFKTGDGLRIAVNDLNLDMFDGQIHGLLGPNGSGKSTTISMLTGLLSPTSGTAF
ILGNDITYQMNEIRKCTGVCLQTDIIWNQLTVLEHLEIYASLKGITNKKQIKSESLKM
ANEIDLGEKLHTPAGSLSGGQKRKLCLGIAFIGRSTIIFLDEVSSGMDPSSRRKVWDF
LLSHKKGKTIILTTHYLDEADYLSDRISIISHGKLITDGSSLFLKNKYGVGYLLTCSK
SLNTIDQFNVDQVTQFIRDQIPDVTVLSNAGSEISFRLPTASLPVFSKFFKDFDENLS
NFHIDSYGISVTTLEEVFLKIGTDIDTDALSIADGNGNDDELKRSIGVSSTGIKSKQQ
LKALLIKRVKTSSKDFKAFLLSLLLPLLVIIGSIIVFKEVDNEVIFYNNSTEPLTFSL
LQQYGKNDIVPIQLAGTTSEFDFNKFLSNSPYFNQFQYLNNSINFNQYLINNYKQSSG
SINFTIPLSSSLINNTTKVISYNSLFNFNYIHSWPVHVNLINDALLRNHNGIGIECTS
LPFDHILTSFQKASQGMNIQSIVYFIVIMMAGFSLMAGSFAGSISQERTNRIKRLLYI
SGCKKYIYWVSNLLWDYFFAFILLLITCIILAIVDDKFNDQFGLFFLSLVLFSLSIIP
LSYLFSFKFSSFGKSTGAITAIHFSIGVIMTIAMIILRIEVIIKNSSSLENIADIIDI
VFNILSPLYAFSRVIFIISGFPGSLRLGALKIDDYWSLHYASTPIIILSVHVVVWTIF
ILLIDYSPEFKGYLRNPKTILPPPPPIDEDSDVSAERIRLESMSPIPSTDQGSGGGGD
MLQYKGLHKLFIGKGKNPNKNAVYNSTLGIPQGQTFGLLGLNGAGKTTTVSMLAGDIL
PTSGQITINGHDLVTDRAQALRGVSVCPQFDALITLLTAREQLSLYCAIKGVPDDKIS
LVVEAFIKMMDLGKIANSNTGGYSGGNKRKVSLSIAMLGNPSVVLIDEASSGCDPIIR
YRQCQVISELGKNKVIILTSHSMSEIQALCSRMTIMRDGQFKCLGSTQHIKSKFGAGY
SVEVKFKKSCLEVGIPQSLQCVLECFPNATILDQHDLMASLELPNPPENPIKVSEIFN
ILSTELSSILDDYSVSQTSLEQVFLKLTGATHEDRLNLNNQQQQQQTIPNSD"

CDS complement(join(7575. .8146,8235. .8364))
/note="ORF_ID:dd_02702"
/codon_start=1
/product="hypothetical protein"
/protein_id="AAM44363.1"
/db_xref="GI:21240652"
/translation="MEIKYFLVLLVGFLLVLPSIVNPYRKGVTITNQQPKINIYQLDV
NNPHNPNNNPHNPHNPNNNPHHPHHLHHHHHHHHHHHHHHHHHHHHHHHHHPHPHH
HPHHHHHPHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHPHPHH
PHPHPHPHPHLHPNPHPHPHPHPHPHPHPHHHPNPNPHPHPHPHHHHHHQEASECL
SYQGNRNRLFIKRDY"

CDS 9039. .10208
/note="ORF_ID:dd_02724"
/codon_start=1
/product="POTASSIUM CHANNEL REGULATORY FACTOR"
/protein_id="AAM44364.1"
/db_xref="GI:21240653"
/translation="MSSLSNLKEQLEKKRKEAQLKNSNNNNNNNNNSNSNNGCDKSE
EPATKKYKTRGEIDKELKQKKEEERIKKEKEKEKEKEKEKEKENYNNNNNSNVNSNKNEN
NINIDNDNDSNDKINKLKENEGFLPDIVVIQRLRERSQPITYFGETSIDRAMRLKKLE
ENDPIEYTQGDNEFANILKQIEQRKMKKHQDSADNEKEKEENENNDKKQQQQQENEE
SEEQELERLRIDSSKCKESKILYFLKGLLNDWITNLENRSDDEKKSRQGKIAEATYYQ
CKSHIQPLFTSLREKTLSIDILEHINEIVEFCKSREYVRANDEYLQMAIGNAPWPMGV
TMVGIHERSSREKIFSNQVAHVLNDEMQRKYIQAVKRLMTFCQTQYPTAPSRSIG"

CDS join(11612. .11806,12045. .12147,12280. .12424,12526. .12693,
12802. .13164,13244. .13607)
/note="ORF_ID:dd_02721"
/codon_start=1
/product="Putative Chromosome segregation ATPase"
/protein_id="AAM44365.1"
/db_xref="GI:21240654"
/translation="MNFRELDTSNGVKAAESSFRTKVVLAREIPWNGFASSNSITSEQ
YNLISKYDKHTDAEKKEKFAANSASYVNFFVNFINSTSNIEIIQYLLTLINEIIEIDP
RAAGAFSKITKDDDKSYPYSVFFRLLNREDAYTNLHASIALAQIMCAGKPTQNDVESF
FNWILKLLRKNNSSEVEVGLIALQSLLLKDDFRIFFNNIDGSALLLNILQALSTSSVN
IQLLYETIYAIWLLTYNKDIAAAYSGTGLVANLVQLVKTVAKEKIVRLSLSTLRNLLN
NGKSNEEMIDNGFVRMLNILNIKKWGDDDIPADIEVLINGLAKDIDNMSSFNKYKTEI
ISGELEWTPVHKSERFWKENISKFEENNYQVIKHLHQTLKTSQSTPLQLSIACHDLCE
FVRHHSRGKAIMTITNQTRYHGYDVKSNEEVKNQALFALQKMMLNNWEYLNAK"

CDS complement(13818. .14342)
/note="ORF_ID:dd_02721"
/codon_start=1
/product="hypothetical protein"
/protein_id="AAM44366.1"
/db_xref="GI:21240655"
/translation="MNENFKIIDKEITEILLKNADFIHHIFKIHYKSLIHNNTDNNNN
DNSGYNEYWNRIIYKDCNNISIPMFIKFKHLFIGNDCLKPINDKDLELLFGQYCFDID
LRNLIINSIPNSNYFNITNSMISKAIDSNRIDYLKYLYENGGEIKNNPSLLQHLKNQL
SKSYDYNCFLDWLN"

CDS complement(14672. .17431)
/note="ORF_ID:dd_02711"
/codon_start=1
/product="hypothetical protein"
/protein_id="AAM44367.1"
/db_xref="GI:21240656"
/translation="MESPNVEILFWKVFHNKYLFSIILKEIHKCEWVNYVYLGYHVDN
RKRFENITSMEWMINNEQFELLKLKINSNDEIIEINKKSIKELFSINDPSLIEQLLTQ
LFQSNKKNQFIKIINEMLPTLPIITSNSLRFKNTYNIFQLLLSNLSSSSFQIYSTL
LECAIEFLQPNTLSILINSPNLFITQDFKNRSLTFTNLNSTSIKTIEEKIELILSNPK
LYNTDYSNLDININFNGNGRERILKFIPKNENFGLSLILLINNSNIQNKLIDLNFHEL
NLTNLVDNKYKINEILLDYNRELRVKELHKYLLLNPTIVNLYDSIYLDHLVYYSRYGE
IIPNLKENNLMDFIKISIATFNVKALKTISKQDLEQFKLKYSKPHSRETTIPLSISRE
YKSKQEMIEFINFYTTNYDYFQCFHFNHNKIIQDYILKNLSKTSSSSSSNIEIEYKEL
LNSINNIEFLNFFPNSWNELIKIIADSNSILLSVEIKDNLQIDWIFENCEKEFLKRSI
RNFKSQEFFHFQNFQLLNYASNLFSSLPRDHYLYFNNNNNQDDDDYTQINSRMVLYSY
PFIKSIRDCNIEQMKYLYSTKGDDFFLYGFNITIRLFAEKQFFFSSNLSDEEIINFIQ
YLNDNQFKFFKPNLVKRVIKSIPDRIKENELLKLKINFNDFIDEASHIFLKSIEKLSI
NQIKLMIGNGFFPIQDISSSLLIRYKDPIYSPLSKCAKFYSISNDNCEKVITLIQIFL
QMYENYTKEEDDDNDNDQDYAFHIIINHLFKVLLSVKKVTIQVIKNIYNQISSSGSSS
SKHIIIDHSLYHVVFYLSIKSPLLARYIMSRPFIYSIFTDDDDGTTPYYNCSDNPQHP
GYDIRKYLNLHVFTFDFILGKDTDTHTSFDLILEKLIDALSYKPMTDDESLLFTEEDY
LNETD"

CDS 18595. .20304
/note="ORF_ID:dd_02710"
/codon_start=1
/product="PAXILLIN-LIKE PROTEIN"
/protein_id="AAM44368.1"
/db_xref="GI:21240657"
/translation="MATKGLNMDDLDLLLADLGRPKSSIKVTATVQTTATPSSGKNFD
NSDIQNEIQSIIEELDQQPQTVQTISTPAPKNHNTTTTTASFSVSSQPAPQPPQQSQQ
IDGLDDLDELMESLNTSISTALKAVPTTPEEHITHANSNSPPPSLHKNTSSTNSASSL
SRPNNNPSVVSTPQPGKVTSTATITTKKQPALSKATLETTSGNNVYSSQPSQSQPQPY
KVTATNSQPSSDDLDELLKGLSPSTTTTTTVPPPVQRDQHQHHHQHQHHHHHNPNHNQ
TQTVTTQINIGRTNTPNNNNNNNTNSPKVVHGDDLDNLLNNLTSQVKDIDSTGPTSRG
TCGGCRKPIFGETIQAMGKFYHPEHFCCHNCQNPLGTKNYYEQESLPHCEKCYQELFC
ARCAHCDEPISDRCITALGKKWHVHHFVCTQCLKPFEGGNFFERDGRPYCEADFYSTF
AVRCGGCNSPIRGECINALGTQWHPEHFVCQYCQKSFTNGQFFEFGGKPYCDVHYHQQ
AGSVCSGCGKAVSGRCVDALDKKWHPEHFVCAFCMNPLAGGSYTANNGKPYCKGCHNK
LFA"

CDS join(21782. .22563,22679. .22793,22936. .27681)
/note="ORF_ID:dd_03308"
/codon_start=1
/product="putative protein contains similarity to
topoisomerase-related function protein"
/protein_id="AAM44369.1"
/db_xref="GI:21240658"
/translation="MDFKKQVLSWVLSIGHEELETILTDCNKDECDKLRAMYSEWKFQ
DGYFFDESVDIPGMPSFDSDSIDDENNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNIGDNNNISNNNINNNNNNINNNNNTINNNNRKLATSFSTSNIINNNSTNNINNSS
SSSSNNNNNNSINGSNTPNNTSINSNCSSNSSSKLKSKKSHSSGSIGNVVCYSDIEMD
RRACSSLLKKFKLEFRLTNDKSIYHFENLQQFLPERKVEEEKRIVKLIRLCDTRTRLD
TFTIESNSVSNQELVHLLTELSQGGFLQVPCHLIKDSQNNWRSEPPKWFISKSFFSLG
EFVAAWFESLIWAQFWENHRIDGRKDQSSKQRNPKPIFKLLDSRKSNLKEFWSKQMTI

```
                    DNRKRYAMAKMGNLLLSVIHSLNETPEKVDLDFIEKLGELIEDFKSPTSTSPSIISSN
                    TTTPTSTTPTTSISTTTTTTTTTSSNNNSNNNNNNNNISSSTPTTPITVSVDKPKLIQ
                    RSTTNRNTNNQSQQNHVSVGGKLGVLGGSSSLSESPENLNSISSIFEEYLVQNPIKLI
                    EFLCFSPLERAGSTVDTVMRQIGAQLQNFYSESLSMDLINSEESSRNRKRIGGNNSNS
                    GSLNSSGNNTPVKQKKKKKQQDQKQQDKEKEKEKEKEKEKEKEKTISPTTTTTTTTT
                    ANSNSLSNSGGGNSSKKLNESSDSTQSDSTNRTKPNANPNTNTSIRKKRKKKSGRGGN
                    NNNNNNNNNQGRGGNSNSNHGGHHHHNHHQSNQGEKSSANSDTEQESIQTRPRRHSLP
                    MDTPPTIKNGIPFLSTTEDEDGVSMLSESDIDHGKGGTGPIKNNNNNNNNNNNNNNNN
                    NNNNNNNNNNNNNNNNNNNNNNSNHSSNKSRNISTSSSTSTSTSNNTNNNSNSSTIG
                    GNSLSSSTTGTSQPLSNITTTKPKQNGKTGTNKKGNSKSSSSTPTKPTVTKNRANSTC
                    EQRTPLILTQKSLTSKSLNNIAVTSTSDSDGIKPKFIYSEEFEEDDEDEGSDEESDED
                    DEDEESDEEYREEDGFNDLEDDEEDQNNSYVQDLTKQIEKHFSLNDEKKNNNNNNNSQ
                    DETLKQPIAKQHQQQQQNQQQQQQVNSSSKLNLNQSLDNIPNANNNNQNGQKVLSPQQ
                    TSQQSNTSPRLFNIFATPNSPNPSTSPILLHQATINSSPKVSSPQPLSPKNGSLSPAG
                    NVSSKRLFFENQTISSGNTSPKITNNNNNSNNINNNNSTTTTTTTNNNNPIATNIQS
                    NINNNEKDKERDNNQKNNNNNNNDNNNNNNNNNINNNNNNTSNNNNNNNNNNINNNNN
                    ISNNNNNNNNNNNNNNNNNNNNINNNNNTMNSSESTNTNTNTTTTTTTTTTTNINNNDR
                    NVLKQSKSTSSLQVTSSVFQPPSKTIKNQKYQYQPPPPQPQQPPIDKYKDWRHNKFFN

  Query Match             4.2%;  Score 51.2;  DB 2;  Length 61052;
  Best Local Similarity   48.0%;  Pred. No. 0.76;
  Matches  146;  Conservative    0;  Mismatches  158;  Indels    0;  Gaps    0;

Qy     886 AGAAAAGAAAGACCAAAACAATGAAAACTTTACGAAAAATCAAAAAGTTGAAGGACTTTA 945
           | |||| ||| |. |||| |||||||||    |  ||  |  |||||    ||  |
Db   28474 AAAAAAAAAAAAAAAAAAAAATGAAAAAAAAAAAAAATGAAAAAAAAAAAAAAATTAAAAAT 28533

Qy     946 ACGTCGAGATCTCTCGTAGAAAACCTCTTTTGTAAGGTTGCATACAATACTTTTTTTTCA 1005
           | |   |  | |     | ||  | |||   ||   |  | |||||||  ||||||
Db   28534 AAATAATAAATTTTGACCCACAAATTATTTGTAAATAATTAAACCAATACTGTTTTTTTT 28593

Qy    1006 GACTTTACTTATGGTATTATACTGAATATGTTATTGCTGTTATAGTAGTTGAGTGACGTT 1065
            |||  | ||   | |    |  ||| | |||  | || |  || ||   || | |.
Db   28594 TTTTTTTTTAATAAAAGTTATATTTTTATTTAATTTTTTTTTTTTTTATTTCTTTGTCCTC 28653

Qy    1066 TGAGGGAATTTCTAGTCCGTTAATCTTGTACTCAGTGTGTCTACTTTTCAAAAAAGTCAG 1125
           | |  |  |   ||  | || |  |   | | | |  |||  ||||||   |
Db   28654 CTCTTTGAATATTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTAAAAAAAAAAAAA 28713

Qy    1126 TTTTTCAGTCTCTAAAACACATTTAAATAAGAGTTTCTTTGCCCATCTTTTGTTCCTCAT 1185
            ||||   | | |   | | |  | || || |    ||||    |  |||| ||  |
Db   28714 ATTTTTTTTGTTTTTTAAAAAAATGAAAAAAATCAACTTTAAAAAATTTTTTTTTTTTTA 28773

Qy    1186 CCTA 1189
           | ||
Db   28774 CATA 28777


RESULT 7
AC116966/c
LOCUS       AC116966               268147 bp    DNA     linear   HTG 04-APR-2002
DEFINITION  Dictyostelium discoideum chromosome 2 map 2482102-2750247 strain
            AX4, *** SEQUENCING IN PROGRESS ***, in ordered pieces.
ACCESSION   AC116966
VERSION     AC116966.1  GI:19920065
KEYWORDS    HTG; HTGS_PHASE2.
SOURCE      Dictyostelium discoideum.
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.
REFERENCE   1  (bases 1 to 268147)
  AUTHORS   Gloeckner,G., Eichinger,L., Szafranski,K., Pachebat,J., Dear,P.,
            Lehmann,R., Baumgart,C., Parra,G., April,J.F., Guigo,R., Kumpf,K.,
            Tunggal,B., Cox,E., Quail,M.A., Platzer,M., Rosenthal,A. and
            Noegel,A.A.
  TITLE     Sequence and Analysis of Chromosome 2 of Dictyostelium
  JOURNAL   Unpublished
  REMARK    The Dictyostelium Genome Sequencing Consortium
REFERENCE   2  (bases 1 to 268147)
  AUTHORS   Baumgart,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-APR-2002) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenberstr. 11, Jena 07745, Germany
COMMENT     CDS predictions from GeneID may contain errors. Further Information
            is available from IMB Jena, Department of Genome Analysis
            (http://genome.imb-jena.de/dictyostelium/)
            and the Univerity Colonge, Institute for Biochemistry I
            (http://www.uni-koeln.de/dictyostelium/project.shtml
            Funding
            Agency : Deutsche Forschungsgemeinschaft (DFG).
            * NOTE: This is a 'working draft' sequence.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
FEATURES             Location/Qualifiers
     source          1. .268147
                     /organism="Dictyostelium discoideum"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /chromosome="2"
                     /map="2482102-2750247"
BASE COUNT   101633 a  30111 c  30461 g 105740 t    202 others
ORIGIN

  Query Match             4.2%;  Score 51.2;  DB 2;  Length 268147;
  Best Local Similarity   43.8%;  Pred. No. 0.58;
  Matches  221;  Conservative    0;  Mismatches  283;  Indels    0;  Gaps    0;

Qy      77 AAAGCAACTTCATCATCATACAGTATGGTTTTGATATGCTCTTCCATTATCACTGAGCCT 136
           | | |||| ||| ||  ||  | | | || ||  ||  |  || ||   |   |     |
Db  254127 ACACCAACATCACCAGTATTAAATGTTGTGTTTGTAGTCATTTTCAAATTTTATTTATTT 254068

Qy     137 TATGATTATGTTTTACGAGCTTATAATATCACTGATGGTGATTCAGTATTGTGATTATGT 196
             | |||       ||    ||| || || | | ||  | ||    |  | | |||| |
Db  254067 ATTTATTGAAAAATAATGTTTTAAAAAATAAATAATTATAATGATTTTATTTTTTTATTT 254008

Qy     197 CCTTCGTTGATTATTCTGTTTCATACAAGTCGTGTAATTTGCTGTTTGTGACAGTACGAT 256
           ||  ||  || || |  || ||     |  |  ||  |    ||  | | | |   ||
Db  254007 ATTTTATTTTTTTTTTTAATTAATTTTTTTTTTTTAAACATATATTTAATTATATTTTAAT 253948

Qy     257 AGATCGACTCAACCTTCTGAGGTATTAGTTGAAGTTCATGTAAATTAGCTTTGTTTATCA 316
           | |    | ||  || | |  | ||    |    ||| | | ||| ||  |  |   |
Db  253947 ATTTTTTTTAAATATTTTAATTTTTTCTCTCTTTTTCTTATTAATCAGTATGTTAAGTTT 253888

Qy     317 TAGTAGCATTTGATTATTGATGCTCTGTAGCTAATGATAAGCCATTGGAGGGAAGCAAGC 376
           || |    ||| |||  |  |            | || |    | ||       |  |
Db  253887 TATTTTTTTTTAATTGATATTTTAGATATTGAATTGCTTTTTCGTTTTTATATATATATG 253828

Qy     377 TTTCTAAATGAATCTACGAATGGATGATAAAGTTCATGAATATTTTTGTTACTTCTGCAG 436
           | |  |||| ||   |  |     | ||     |       || |       |  |
Db  253827 TATAAAAATAAAAAAAATATAATTTTATCTTTGTTGGTTTTAATGAATCATTTATTATTA 253768

Qy     437 TCAGATCATGAGTTATTGAGTCTATTGTTTTTTTAAGCCTGTTTCAGATGATCCATCATC 496
           | |  |     || || | | | || | |||||       ||| | ||  |   | ||
Db  253767 TTATTTTTAATATTTTTTACTTTTTTTTTGTTTTTTTATAATTTTTATATATTTATTAATT 253708

Qy     497 AGTAACAACATACACGGTGTAGTCCCAAATCCATCATATGCACCTTCTTTTCTTCAATTT 556
           | ||| |  ||      | |  |     ||  || || |  |  || | || ||   |||
Db  253707 ATTAATATTATTATTATTATTATTAATTATTTATTATTTATAATTTTTATTTTTATTTTT 253648

Qy     557 GGTCTTGTTTTTTTTTTTTCATGA 580
           | || ||||| ||||||  | |
Db  253647 ATTTTTATTTTTATTTTTTTTTTA 253624


RESULT 8
AX345543
LOCUS       AX345543               6531 bp    DNA     linear   PAT 01-FEB-2002
DEFINITION  Sequence 614 from Patent WO0200928.
ACCESSION   AX345543
VERSION     AX345543.1  GI:18493429
KEYWORDS    .
SOURCE      synthetic construct.
  ORGANISM  synthetic construct
            artificial sequences.
REFERENCE   1
  AUTHORS   Olek,A., Piepenbrock,C. and Berlin,K.
```

```
  TITLE     Diagnosis of diseases associated with the immune system
  JOURNAL   Patent: WO 0200928-A 614 03-JAN-2002;
            Epigenomics AG (DE)
FEATURES             Location/Qualifiers
     source          1. .6531
                     /organism="synthetic construct"
                     /db_xref="taxon:32630"
                     /note="chemically treated genomic DNA (Homo sapiens)"
BASE COUNT     1742 a    131 c   1301 g   3357 t
ORIGIN

  Query Match             4.2%;  Score 51;  DB 6;  Length 6531;
  Best Local Similarity   44.2%;  Pred. No. 1.3;
  Matches  210;  Conservative    0;  Mismatches  265;  Indels    0;  Gaps    0;

Qy     157 TTATAATATCACTGATGGTGATTCAGTATTGTGATTATGTCCTTCGTTGATTATTCTGTT 216
           |||| | |   | | |||||||  | ||| |     |    |   ||  ||||| ||
Db    5002 TTATTTTGTTGTTTAGGGTGATTTTGAATTTTTGCGTTTAAGTAGTTTTTTTATTTTGGA 5061

Qy     217 TCATACAAGTCGTGTAATTTGCTGTTTGTGACAGTACGATAGATCGACTCAACCTTCTGA 276
           |  |  ||||  ||  |||   ||||| |         |   | | |     | | |
Db    5062 TTTTTAAAGTGTTGGGATTATAAGTTTGAGTTTTATTTTTTTTTTTAATTGTTTTAGTTA 5121

Qy     277 GGTATTAGTTGAAGTTCATGTAAATTAGCTTTGTTTATCATAGTAGCATTTGATTATTGA 336
           | ||| ||    ||  | |   ||   ||| |||    ||||      | |  ||| |
Db    5122 TTTTTTAATTTTTTTTTTTTTTTTTTTTTTTTTTTGATTTAGTTATTGGTAAAAATTTA 5181

Qy     337 TGCTCTGTAGCTAATGATAAGCCATTGGAGGGAAGCAAGCTTTCTAAATGAATCTACGAA 396
           | | | |      || | |   || ||||   |   || ||| ||  | ||     ||
Db    5182 GGTTTTTTTTTATTTGTTTATTTATAGGAGAATAATTAGTTTTTTATTTTAAGGAGGAAA 5241

Qy     397 TGGATGATAAAGTTCATGAATATTTTTGTTACTTCTGCAGTCAGATCATGAGTTATTGAG 456
           | |      | || |    | ||||| ||     |  |||   ||  | | ||  |
Db    5242 TAGTATGATTATTTTAATTTTTTTTTTTTTTAAGTTTTAGTTTAATAGTAATTTTATTTA 5301

Qy     457 TCTATTGTTTTTTTAAGCCTGTTTCAGATGATCCATCATCAGTAACAACATACACGGTGT 516
           | | |  |||||||     ||||| |||      | |   ||      |  | |||
Db    5302 TTTTTATTTTTTTTTTATTTGTTTTAGAAATAAAAGTAATTTTATTTTTTTTTAAGGTAA 5361

Qy     517 AGTCCCAAATCCATCATATGCACCTTCTTTTCTTCAATTTGGTCTTGTTTTTTTTTTTTC 576
           | |      |   | |        || ||||     ||| |  | |  |||  || |
Db    5362 ATTTTTGTTTAGGTTAGTATGTTGTTTTTTTTAAATTTTTTGAGATTTGGTTTAATTATT 5421

Qy     577 ATGATGTCATTGAATTATTCAAGAAGTCACTTCGAGCATAATGATTTTTCAAAAT 631
           | || |  ||  | ||  ||   | | ||   |  | |  || ||   |||
Db    5422 TTTATTTTTTTTTTGTATTTTTAGTTTTTAGTTTTTGAGGATTTTTTATTTTTAAT 5476


RESULT 9
AC008929/c
LOCUS       AC008929               125403 bp    DNA     linear   PRI 23-JAN-2002
DEFINITION  Homo sapiens chromosome 5, complete sequence.
ACCESSION   AC008929
VERSION     AC008929.3  GI:6598820
KEYWORDS    HTG.
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 125403)
  AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 125403)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE   3  (bases 1 to 125403)
  AUTHORS   DOE Joint Genome Center and Stanford Human Genome Center.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-DEC-1999) DOE Joint Genome Institute, 2800 Mitchell
            Drive, Walnut Creek, CA 94598, USA
REFERENCE   4  (bases 1 to 125403)
  AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-FEB-2000) DOE Joint Genome Institute, 2800 Mitchell
            Drive, Walnut Creek, CA 94598, USA
REFERENCE   5  (bases 1 to 125403)
  AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-FEB-2000) DOE Joint Genome Institute, 2800 Mitchell
            Drive, Walnut Creek, CA 94598, USA
REFERENCE   6  (bases 1 to 125403)
  AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-APR-2000) DOE Joint Genome Institute, 2800 Mitchell
            Drive, Walnut Creek, CA 94598, USA
REFERENCE   7  (bases 1 to 125403)
  AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JAN-2002) DOE Joint Genome Institute, 2800 Mitchell
            Drive, Walnut Creek, CA 94598, USA
COMMENT     On Dec 17, 1999 this sequence version replaced gi:6165130.
            Draft Sequence Produced by DOE Joint Genome Institute
            www.jgi.doe.gov
            Finishing Completed at Stanford Human Genome Center
            www-shgc.stanford.edu
            Quality: Phrap Quality >=40 99.8% of Sequence;
            Estimated Total Number of Errors is 0.3.
FEATURES             Location/Qualifiers
     source          1. .125403
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="5"
BASE COUNT     39991 a   24654 c   23616 g   37142 t
ORIGIN

  Query Match             4.2%;  Score 51;  DB 9;  Length 125403;
  Best Local Similarity   52.6%;  Pred. No. 0.73;
  Matches  111;  Conservative    0;  Mismatches  100;  Indels    0;  Gaps    0;

Qy     968 ACCTCTTTTGTAAGGTTGCATACAATACTTTTTTTTCAGACTTTACTTATGGTATTATAC 1027
           | || |||| |||   |     || |||||||| || ||||  || | | ||||  | |||
Db   71354 AGCTTTTTTTTAAATTCAGTGACTATACTTTTATTCCAGAAGTTCCATTTGGTCCTTTAC 71295

Qy    1028 TGAATATGTTATTGCTGTTATAGTAGTTGAGTGACGTTTGAGGGAATTTCTAGTCCGTTA 1087
           |||  ||   || ||||| |  || ||   |    ||          |||||  || |
Db   71294 AAAATTTGAGTTTCCTGTTCTCATATTTTTCTTTTCTTGTCTATTTGTTCTACACCTTCC 71235

Qy    1088 ATCTTGTACTCAGTGTGTCTACTTTTCAAAAAAGTCAGTTTTTCAGTCTCTAAAACACAT 1147
           || | |  |||      |  | ||||    ||||||  ||||  | | | |||       |
Db   71234 TTCATCTTATCAAACATTTAAATTTTTCTTAAAGTCTCTTTTAGAATGTTTAATTTTTTT 71175

Qy    1148 TTAAATAAGAGTTTCTTTGCCCATCTTTTGT 1178
              |  | | | |  |  ||  | |||||
Db   71174 AGTTCTTGGGGGTACAATTGCCCCCATTTGT 71144


RESULT 10
AC027347
LOCUS       AC027347               148193 bp    DNA     linear   PRI 21-JUN-2001
DEFINITION  Homo sapiens chromosome 5 clone CTD-2562I21, complete sequence.
ACCESSION   AC027347
VERSION     AC027347.5  GI:14518405
KEYWORDS    HTG.
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 148193)
  AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 148193)
  AUTHORS   DOE Joint Genome Institute.
```

```
  TITLE     Direct Submission
  JOURNAL   Submitted (30-MAR-2000) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE   3  (bases 1 to 148193)
  AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-JUN-2001) DOE Joint Genome Institute, 2800 Mitchell
            Drive, Walnut Creek, CA 94598, USA
COMMENT     On Jun 21, 2001 this sequence version replaced gi:13699717.
            Draft Sequence Produced by DOE Joint Genome Institute
            www.jgi.doe.gov
            Finishing Completed at Stanford Human Genome Center
            www-shgc.stanford.edu
            Quality: Phrap Quality >=40 99.9% of Sequence;
            Estimated Total Number of Errors is 0.1.
FEATURES             Location/Qualifiers
     source          1. .148193
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="5"
                     /clone="CTD-2562I21"
BASE COUNT    45847 a  28027 c  28848 g  45471 t
ORIGIN

  Query Match             4.2%;  Score 51;  DB 9;  Length 148193;
  Best Local Similarity   52.6%;  Pred. No. 0.71;
  Matches  111;  Conservative    0;  Mismatches  100;  Indels    0;  Gaps    0;

Qy     968 ACCTCTTTTGTAAGGTTGCATACAATACTTTTTTTTCAGACTTTACTTATGGTATTATAC 1027
           | || |||| |||  |     || |||||||| || ||||  || | | ||||  | |||
Db   13180 AGCTTTTTTTTAAATTCAGTGACTATACTTTTATTCCAGAAGTTCCATTTGGTCCTTTAC 13239

Qy    1028 TGAATATGTTATTGCTGTTATAGTAGTTGAGTGACGTTTGAGGGAATTTCTAGTCCGTTA 1087
             ||| ||   || ||||| |  || ||   |    ||          |||||  || |
Db   13240 AAAATTTGAGTTTCCTGTTCTCATATTTTTCTTTTCTTGTCTATTTGTTCTACACCTTCC 13299

Qy    1088 ATCTTGTACTCAGTGTGTCTACTTTTCAAAAAAGTCAGTTTTTCAGTCTCTAAAACACAT 1147
            || | |  |||     |  | ||||    ||||||  ||||  | | | |||      |
Db   13300 TTCATCTTATCAAACATTTAAATTTTTCTTAAAGTCTCTTTTAGAATGTTTAATTTTTTT 13359

Qy    1148 TTAAATAAGAGTTTCTTTGCCCATCTTTTGT 1178
             | | | | |  |  ||  | |||||
Db   13360 AGTTCTTGGGGGTACAATTGCCCCCATTTGT 13390


RESULT 11
AC034179
LOCUS       AC034179               220146 bp    DNA     linear   HTG 12-NOV-2000
DEFINITION  Homo sapiens chromosome 5 clone RP11-797O11, WORKING DRAFT
            SEQUENCE, 20 unordered pieces.
ACCESSION   AC034179
VERSION     AC034179.3  GI:11139352
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 220146)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Homo sapiens clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 220146)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-APR-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     On Nov 12, 2000 this sequence version replaced gi:7524008.

            ---------- Genome Center ----------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu/gsc/index.shtml

            ---------- Project Information ----------
            Center project name: H_NH0797O11
            ---------- Summary Statistics ----------
            Sequencing vector: M13; 95%
            Sequencing vector: plasmid; 5%
            Chemistry: Dye-primer ET; 95% of reads
            Chemistry: Dye-terminator Big Dye; 5% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 210550 bases at least Q40
            Consensus quality: 213061 bases at least Q30
            Consensus quality: 214675 bases at least Q20
            Insert size: 205000; agarose-fp
            Insert size: 218246; sum-of-contigs
            Quality coverage: 4.85 in Q20 bases; agarose-fp
            Quality coverage: 4.64 in Q20 bases; sum-of-contigs
            ----------------------------------------------------
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 20 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *      1     6923: contig of 6923 bp in length
            *   6924     7023: gap of unknown length
            *   7024    10977: contig of 3954 bp in length
            *  10978    11077: gap of unknown length
            *  11078    17882: contig of 6805 bp in length
            *  17883    17982: gap of unknown length
            *  17983    22351: contig of 4369 bp in length
            *  22352    22451: gap of unknown length
            *  22452    28667: contig of 6216 bp in length
            *  28668    28767: gap of unknown length
            *  28768    35913: contig of 7146 bp in length
            *  35914    36013: gap of unknown length
            *  36014    43927: contig of 7914 bp in length
            *  43928    44027: gap of unknown length
            *  44028    51182: contig of 7155 bp in length
            *  51183    51282: gap of unknown length
            *  51283    63340: contig of 12058 bp in length
            *  63341    63440: gap of unknown length
            *  63441    85661: contig of 22221 bp in length
            *  85662    85761: gap of unknown length
            *  85762   105434: contig of 19673 bp in length
            * 105435   105534: gap of unknown length
            * 105535   134835: contig of 29301 bp in length
            * 134836   134935: gap of unknown length
            * 134936   198011: contig of 63076 bp in length
            * 198012   198111: gap of unknown length
            * 198112   200255: contig of 2144 bp in length
            * 200256   200355: gap of unknown length
            * 200356   203493: contig of 3138 bp in length
            * 203494   203593: gap of unknown length
            * 203594   205664: contig of 2071 bp in length
            * 205665   205764: gap of unknown length
            * 205765   208649: contig of 2885 bp in length
            * 208650   208749: gap of unknown length
            * 208750   212298: contig of 3549 bp in length
            * 212299   212398: gap of unknown length
            * 212399   216138: contig of 3740 bp in length
            * 216139   216238: gap of unknown length
            * 216239   220146: contig of 3908 bp in length.
FEATURES             Location/Qualifiers
     source          1. .220146
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="5"
                     /clone="RP11-797O11"
     misc_feature    1. .6923
                     /note="assembly_name:Contig10"
     misc_feature    7024. .10977
                     /note="assembly_name:Contig11"
```

```
     misc_feature    11078. .17882
                     /note="assembly_name:Contig12"
     misc_feature    17983. .22351
                     /note="assembly_name:Contig13"
     misc_feature    22452. .28667
                     /note="assembly_name:Contig14"
     misc_feature    28768. .35913
                     /note="assembly_name:Contig15"
     misc_feature    36014. .43927
                     /note="assembly_name:Contig16"
     misc_feature    44028. .51182
                     /note="assembly_name:Contig17"
     misc_feature    51283. .63340
                     /note="assembly_name:Contig18"
     misc_feature    63441. .85661
                     /note="assembly_name:Contig19"
     misc_feature    85762. .105434
                     /note="assembly_name:Contig20"
     misc_feature    105535. .134835
                     /note="assembly_name:Contig21
                     clone_end:SP6
                     vector_side:right"
     misc_feature    134936. .198011
                     /note="assembly_name:Contig22"
     misc_feature    198112. .200255
                     /note="assembly_name:Contig3"
     misc_feature    200356. .203493
                     /note="assembly_name:Contig4"
     misc_feature    203594. .205664
                     /note="assembly_name:Contig5"
     misc_feature    205765. .208649
                     /note="assembly_name:Contig6"
     misc_feature    208750. .212298
                     /note="assembly_name:Contig7"
     misc_feature    212399. .216138
                     /note="assembly_name:Contig8"
     misc_feature    216239. .220146
                     /note="assembly_name:Contig9"
BASE COUNT    66033 a  41274 c  41775 g  69151 t   1913 others
ORIGIN

  Query Match             4.2%;  Score 51;  DB 2;  Length 220146;
  Best Local Similarity   52.6%;  Pred. No. 0.66;
  Matches  111;  Conservative    0;  Mismatches  100;  Indels    0;  Gaps    0;

Qy     968 ACCTCTTTTGTAAGGTTGCATACAATACTTTTTTTTTCAGACTTTACTTATGGTATTATAC 1027
           | || |||| |||  |     || |||||||| || ||||  || | | ||||  | |||
Db  200529 AGCTTTTTTTTAAATTCAGTGACTATACTTTTATTCCAGAAGTTCCATTTGGTCCTTTAC 200588

Qy    1028 TGAATATGTTATTGCTGTTATAGTAGTTGAGTGACGTTTGAGGGAATTTCTAGTCCGTTA 1087
             ||| ||   || ||||| |  || ||   |    ||          |||||  || |
Db  200589 AAAATTTGAGTTTCCTGTTCTCATATTTTTCTTTTCTTGTCTATTTGTTCTACACCTTCC 200648

Qy    1088 ATCTTGTACTCAGTGTGTCTACTTTTCAAAAAAGTCAGTTTTTCAGTCTCTAAAACACAT 1147
            || | |  |||     |  | ||||    ||||||  ||||  | | | |||       |
Db  200649 TTCATCTTATCAAACATTTAAATTTTTCTTAAAGTCTCTTTTAGAATGTTTAATTTTTTT 200708

Qy    1148 TTAAATAAGAGTTTCTTTGCCCATCTTTTGT 1178
               |  | | | |  |  ||  | |||||
Db  200709 AGTTCTTGGGGGTACAATTGCCCCCATTTGT 200739


RESULT 12
AC116365/c
LOCUS       AC116365               223228 bp    DNA     linear   HTG 27-MAR-2002
DEFINITION  Homo sapiens chromosome 5 clone RP11-797O11, WORKING DRAFT
            SEQUENCE, 13 unordered pieces.
ACCESSION   AC116365
VERSION     AC116365.1  GI:19745047
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 223228)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Sequencing of Human Chromosome 5
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 223228)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-MAR-2002) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT     -------Genome Center
            Center: Joint Genome Institute
            Center Code: JGI
            Web site: http://www.jgi.doe.gov
            -------
            Project Information
            Center Project Name: 1600899
            Center clone name: RPCI-11_797O11
            -------
            Summary Statistics
            Consensus quality: 209063 bases at least Q40
            Consensus quality: 214412 bases at least Q30
            Consensus quality: 217045 bases at least Q20
            Estimated insert size: 175000; agarose-fp estimation
            Estimated insert size: 222028; sum-of-contigs estimation
            Quality coverage:  6.11 in Q20 bases; agarose-fp estimation
            Quality coverage:  4.81 in Q20 bases; sum-of-contigs estimation.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 13 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     3678: contig of 3678 bp in length
            *     3679     3778: gap of unknown length
            *     3779     9781: contig of 6003 bp in length
            *     9782     9881: gap of unknown length
            *     9882    16585: contig of 6704 bp in length
            *    16586    16685: gap of unknown length
            *    16686    23898: contig of 7213 bp in length
            *    23899    23998: gap of unknown length
            *    23999    34628: contig of 10630 bp in length
            *    34629    34728: gap of unknown length
            *    34729    47189: contig of 12461 bp in length
            *    47190    47289: gap of unknown length
            *    47290    60312: contig of 13023 bp in length
            *    60313    60412: gap of unknown length
            *    60413    73825: contig of 13413 bp in length
            *    73826    73925: gap of unknown length
            *    73926    87955: contig of 14030 bp in length
            *    87956    88055: gap of unknown length
            *    88056   102751: contig of 14696 bp in length
            *   102752   102851: gap of unknown length
            *   102852   125755: contig of 22904 bp in length
            *   125756   125855: gap of unknown length
            *   125856   163416: contig of 37561 bp in length
            *   163417   163516: gap of unknown length
            *   163517   223228: contig of 59712 bp in length.
FEATURES             Location/Qualifiers
     source          1. .223228
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="5"
                     /clone="RP11-797O11"
                     /clone_lib="RPCI human BAC library 11"
BASE COUNT    70216 a  42507 c  42023 g  67274 t   1208 others
ORIGIN

  Query Match             4.2%;  Score 51;  DB 2;  Length 223228;
  Best Local Similarity   52.6%;  Pred. No. 0.66;
  Matches  111;  Conservative    0;  Mismatches  100;  Indels    0;  Gaps    0;
```

```
Qy     968 ACCTCTTTTGTAAGGTTGCATACAATACTTTTTTTTCAGACTTTACTTATGGTATTATAC 1027
           | || |||| |||  |     || |||||||| || ||||  || | | ||||  | |||
Db   91014 AGCTTTTTTTTAAATTCAGTGACTATACTTTTATTCCAGAAGTTCCATTTGGTCCTTTAC 90955

Qy    1028 TGAATATGTTATTGCTGTTATAGTAGTTGAGTGACGTTTGAGGGAATTTCTAGTCCGTTA 1087
             ||| ||   || ||||| |  || ||   |    ||          |||||  || |
Db   90954 AAAATTTGAGTTTCCTGTTCTCATATTTTTCTTTTCTTGTCTATTTGTTCTACACCTTCC 90895

Qy    1088 ATCTTGTACTCAGTGTGTCTACTTTTCAAAAAAGTCAGTTTTTCAGTCTCTAAAACACAT 1147
            || | |  |||      |  | ||||    ||||||  ||||  | | | |||      |
Db   90894 TTCATCTTATCAAACATTTAAATTTTTCTTAAAGTCTCTTTTAGAATGTTTAATTTTTTT 90835

Qy    1148 TTAAATAAGAGTTTCTTTGCCCATCTTTTGT 1178
              |  | | | |  |  ||  | |||||
Db   90834 AGTTCTTGGGGGTACAATTGCCCCCATTTGT 90804
```

RESULT 13
HS1012F16/c
LOCUS HS1012F16 74539 bp DNA linear PRI 15-MAR-2001
DEFINITION Human DNA sequence from clone RP5-1012F16 on chromosome 20. Contains a novel gene, ESTs, STSs and GSSs, complete sequence.
ACCESSION AL080274
VERSION AL080274.21 GI:7594620
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 74539)
AUTHORS Phillimore,B.
TITLE Direct Submission
JOURNAL Submitted (08-MAR-2001) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT On Apr 18, 2000 this sequence version replaced gi:6630760.
During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.
The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 20, constructed by the Sanger Centre Chromosome 20 Mapping Group. Further information can be found at http://www.sanger.ac.uk/HGP/Chr20
IMPORTANT: This sequence is not the entire insert of clone RP5-1012F16 It may be shorter because we sequence overlapping sections only once, except for a 100 base overlap.
The true left end of clone RP5-1012F16 is at 1 in this sequence.
The true left end of clone RP4-742J24 is at 74440 in this sequence.
This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. RP5-1012F16 is from the library RPCI-5 constructed by the group of Pieter de Jong. For further details see http://www.chori.org/bacpac/home.htm
VECTOR: pCYPAC2.

FEATURES Location/Qualifiers

source	1. .74539
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/chromosome="20"
	/clone="RP5-1012F16"
	/clone_lib="RPCI-5"
repeat_region	68. .375
	/note="AluSx repeat: matches 1. .308 of consensus"
gene	complement(727. .40353)
	/gene="dJ1012F16.1"
mRNA	complement(join(727. .1181,40249. .>40353))
	/gene="dJ1012F16.1"
	/product="dJ1012F16.1 (novel protein)"
	/note="match: ESTs: Em:AA405155 Em:AA405616 Em:AI200948 Em:AW592501 Em:AW263234"
	/evidence=not_experimental
polyA_site	complement(727)
	/gene="dJ1012F16.1"
repeat_region	728. .934
	/note="LTR16C repeat: matches 141. .344 of consensus"
misc_feature	729. .1072
	/note="match: STS: Em:AA405616"
polyA_signal	complement(744. .749)
	/gene="dJ1012F16.1"
CDS	complement(join(974. .1181,40249. .>40353))
	/gene="dJ1012F16.1"
	/codon_start=2
	/evidence=not_experimental
	/product="dJ1012F16.1 (novel protein)"
	/protein_id="CAB88097.1"
	/db_xref="GI:7594621"
	/db_xref="SPTREMBL:Q9NU29"
	/translation="RKERRKGGRKENREPRERREKIRKKKEREKGRTWILFVACAEDY ISLLQLSSTLKNKEDRGEASGPHHPLQADCFLLQALEHLPESFLGQWEPGLAQSQAGE V"
repeat_region	1212. .1273
	/note="2 copies 31 mer 100% conserved"
repeat_region	2101. .2402
	/note="AluY repeat: matches 1. .302 of consensus"
repeat_region	3026. .3124
	/note="L2 repeat: matches 2539. .2632 of consensus"
repeat_region	3126. .3168
	/note="MER1B repeat: matches 294. .337 of consensus"
repeat_region	3169. .3788
	/note="MER51B repeat: matches 1. .617 of consensus"
repeat_region	3789. .4094
	/note="MER1B repeat: matches 1. .294 of consensus"
repeat_region	4163. .4620
	/note="MLT1C repeat: matches 1. .466 of consensus"
misc_feature	4732. .5095
	/note="match: GSS: Em:B94662"
misc_feature	4737. .5111
	/note="match: GSS: Em:AQ085534"
repeat_region	4748. .5042
	/note="LTR16C repeat: matches 79. .387 of consensus"
misc_feature	complement(5656. .6114)
	/gene="dJ1012F16.1"
	/note="match: GSS: Em:AQ778148"
repeat_region	5818. .5926
	/note="MIR repeat: matches 110. .233 of consensus"
repeat_region	6327. .6369
	/note="MLT1G repeat: matches 196. .236 of consensus"
repeat_region	6611. .6670
	/note="MLT1C repeat: matches 452. .523 of consensus"
repeat_region	7507. .7654
	/note="L2 repeat: matches 2260. .2407 of consensus"
repeat_region	7667. .7710
	/note="22 copies 2 mer tc 100% conserved"
repeat_region	7797. .7981
	/note="LTR16C repeat: matches 189. .379 of consensus"
repeat_region	8290. .8334
	/note="L2 repeat: matches 2440. .2484 of consensus"
repeat_region	8594. .9430
	/note="L1MD2 repeat: matches 5290. .6122 of consensus"
repeat_region	9488. .9549
	/note="31 copies 2 mer at 69% conserved"
repeat_region	9506. .9549

```
                     /note="11 copies 4 mer agat 90% conserved"
repeat_region        9519. .9580
                     /note="2 copies 31 mer 95% conserved"
repeat_region        9550. .9593
                     /note="11 copies 4 mer gata 97% conserved"
repeat_region        9597. .9832
                     /note="7SK repeat: matches 1. .193 of consensus"
repeat_region        9841. .9876
                     /note="9 copies 4 mer gata 94% conserved"
repeat_region        9845. .9906
                     /note="2 copies 31 mer 98% conserved"
repeat_region        9908. .10019
                     /note="L1MD3 repeat: matches 7631. .7739 of consensus"
repeat_region        12316. .12615
                     /note="AluJo repeat: matches 1. .291 of consensus"
repeat_region        12631. .12670
                     /note="20 copies 2 mer ac 100% conserved"
repeat_region        12821. .12860
                     /note="20 copies 2 mer gt 100% conserved"
misc_feature         13109. .13501
                     /note="match: GSS: Em:AQ819000"
repeat_region        14148. .14534
                     /note="MER67C repeat: matches 6. .403 of consensus"
repeat_region        14597. .14671
                     /note="MER67D repeat: matches 433. .508 of consensus"
repeat_region        15114. .15626
                     /note="MER82 repeat: matches 1. .516 of consensus"
repeat_region        15627. .15914
                     /note="AluY repeat: matches 1. .288 of consensus"
repeat_region        15915. .16053
                     /note="MER82 repeat: matches 515. .653 of consensus"
repeat_region        16303. .16367
                     /note="Charlie4 repeat: matches 1871. .1936 of consensus"
repeat_region        17073. .17152
                     /note="L1MC4 repeat: matches 7553. .7638 of consensus"
repeat_region        17153. .17494
                     /note="L1PA16 repeat: matches 5812. .6157 of consensus"
repeat_region        17495. .17878
                     /note="L1MC4 repeat: matches 7638. .7977 of consensus"
repeat_region        18086. .18137
                     /note="26 copies 2 mer tc 84% conserved"
repeat_region        18088. .18139
                     /note="13 copies 4 mer tctc 84% conserved"
repeat_region        18163. .18332
                     /note="MLT1G repeat: matches 115. .285 of consensus"
repeat_region        18383. .18732
                     /note="MLT1A1 repeat: matches 2. .364 of consensus"
repeat_region        19124. .19256
                     /note="LTR16C repeat: matches 254. .387 of consensus"
misc_feature         19237. .19593
                     /note="match: STS: Em:AL031203 Em:HS901B8T"
repeat_region        19640. .20041
                     /note="MER63B repeat: matches 4. .401 of consensus"
misc_feature         complement(20472. .20886)
                     /gene="dJ1012F16.1"
                     /note="match: GSS: Em:AQ088743"
misc_feature         20944. .21448
                     /note="match: GSS: Em:AQ476653"
repeat_region        21365. .21449
                     /note="MLT1I repeat: matches 58. .144 of consensus"
repeat_region        21504. .21763
                     /note="MLT1H repeat: matches 14. .286 of consensus"
misc_feature         21727. .22163
                     /note="match: GSS: Em:AQ451319"
repeat_region        22003. .22175
                     /note="MLT1I repeat: matches 232. .410 of consensus"
repeat_region        22385. .22547
                     /note="FAM repeat: matches 1. .166 of consensus"
repeat_region        23254. .23425
                     /note="L1ME3 repeat: matches 5521. .5694 of consensus"
repeat_region        23426. .23709
                     /note="AluJo repeat: matches 5. .290 of consensus"
repeat_region        23710. .24130
                     /note="L1ME3 repeat: matches 5694. .6104 of consensus"

  Query Match             4.1%;  Score 49.8;  DB 9;  Length 74539;
  Best Local Similarity   50.2%;  Pred. No. 1.4;
  Matches  123;  Conservative    0;  Mismatches  122;  Indels    0;  Gaps    0;

Qy     934 TGAAGGACTTTAACGTCGAGATCTCTCGTAGAAAACCTCTTTTGTAAGGTTGCATACAAT 993
           ||   ||||||||  |  |||    || |||  |   |   ||  ||  || | |  | |
Db   46547 TGTCTGACTTTAATTTGAAGAAAATTCTTAGGCATTATTGCTTTAAATATTTCTTTTATT 46488

Qy     994 ACTTTTTTTTCAGACTTTACTTATGGTATTATACTGAATATGTTATTGCTGTTATAGTAG 1053
           ||||  | |    | | | ||| || ||||    | |     || ||    || | ||||
Db   46487 TCTTTCTCTCTTTATTCTCCTTCTGCTATTCCCATTACACATTTGTTATAATTTTTGTAG 46428

Qy    1054 TTGAGTGACGTTTGAGGGAATTTCTAGTCCGTTAATCTTGTACTCAGTGTGTCTACTTTT 1113
           |||  | | |  ||   |||  |||||||||  |  | |||   |   | | | | |||||
Db   46427 TTGCTTCAGGGTTCTTGGATATTCTAGTCCTCTTTTTTTGGTATTTTTTCTTTGCTTTT 46368

Qy    1114 CAAAAAAGTCAGTTTTTCAGTCTCTAAAACACATTTAAATAAGAGTTTCTTTGCCCATCT 1173
           ||     |   |||| |       ||  |   |  |    ||   |||| |    ||| |
Db   46367 CAGTTTTGGATGTTTATATTGACATATCATCAAGCTTGGAAATTCTTTCCTCAGTCATAT 46308

Qy    1174 TTTGT 1178
              ||
Db   46307 CCAGT 46303


RESULT 14
AC008375
LOCUS       AC008375               127811 bp    DNA     linear   HTG 06-MAY-2000
DEFINITION  Homo sapiens chromosome 19 clone CTC-203B18, WORKING DRAFT
            SEQUENCE, 27 ordered pieces.
ACCESSION   AC008375
VERSION     AC008375.6  GI:7711254
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT.
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 127811)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Sequencing of Human Chromosome 19
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 127811)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT     On May 6, 2000 this sequence version replaced gi:7689753.
            -------Genome Center
            Center: Joint Genome Institute
            Center Code: JGI
            Web site: http://www.jgi.doe.gov
            -------
            Project Information
            Center Project Name: 263173, BC228196
            Center clone name: CIT-HSPC_203B18
            -------
            Summary Statistics
            Consensus quality: 108189 bases at least Q40
            Consensus quality: 121159 bases at least Q30
            Consensus quality: 123401 bases at least Q20
            Estimated insert size: 101730; agarose-fp estimation
            Estimated insert size: 125211; sum-of-contigs estimation
            Quality coverage:  6.7 in Q20 bases; agarose-fp estimation
            Quality coverage:  5.38 in Q20 bases; sum-of-contigs estimation.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 27 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
```

```
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *       1     4137: contig of 4137 bp in length
            *    4138     4237: gap of unknown length
            *    4238     6184: contig of 1947 bp in length
            *    6185     6284: gap of unknown length
            *    6285     7190: contig of 906 bp in length
            *    7191     7290: gap of unknown length
            *    7291    11893: contig of 4603 bp in length
            *   11894    11993: gap of unknown length
            *   11994    17297: contig of 5304 bp in length
            *   17298    17397: gap of unknown length
            *   17398    20788: contig of 3391 bp in length
            *   20789    20888: gap of unknown length
            *   20889    27303: contig of 6415 bp in length
            *   27304    27403: gap of unknown length
            *   27404    34878: contig of 7475 bp in length
            *   34879    34978: gap of unknown length
            *   34979    45862: contig of 10884 bp in length
            *   45863    45962: gap of unknown length
            *   45963    47282: contig of 1320 bp in length
            *   47283    47382: gap of unknown length
            *   47383    58063: contig of 10681 bp in length
            *   58064    58163: gap of unknown length
            *   58164    70353: contig of 12190 bp in length
            *   70354    70453: gap of unknown length
            *   70454    77902: contig of 7449 bp in length
            *   77903    78002: gap of unknown length
            *   78003    80724: contig of 2722 bp in length
            *   80725    80824: gap of unknown length
            *   80825    84590: contig of 3766 bp in length
            *   84591    84690: gap of unknown length
            *   84691    87850: contig of 3160 bp in length
            *   87851    87950: gap of unknown length
            *   87951    93467: contig of 5517 bp in length
            *   93468    93567: gap of unknown length
            *   93568    99623: contig of 6056 bp in length
            *   99624    99723: gap of unknown length
            *   99724   102265: contig of 2542 bp in length
            *  102266   102365: gap of unknown length
            *  102366   106161: contig of 3796 bp in length
            *  106162   106261: gap of unknown length
            *  106262   113210: contig of 6949 bp in length
            *  113211   113310: gap of unknown length
            *  113311   115651: contig of 2341 bp in length
            *  115652   115751: gap of unknown length
            *  115752   117985: contig of 2234 bp in length
            *  117986   118085: gap of unknown length
            *  118086   122100: contig of 4015 bp in length
            *  122101   122200: gap of unknown length
            *  122201   124814: contig of 2614 bp in length
            *  124815   124914: gap of unknown length
            *  124915   127320: contig of 2406 bp in length
            *  127321   127420: gap of unknown length
            *  127421   127811: contig of 391 bp in length.
FEATURES             Location/Qualifiers
     source          1. .127811
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="19"
                     /clone="CTC-203B18"
                     /clone_lib="CalTech human BAC library C"
BASE COUNT    36523 a  25019 c  25153 g  38299 t   2817 others
ORIGIN

  Query Match             4.1%;  Score 49.8;  DB 2;  Length 127811;
  Best Local Similarity   54.7%;  Pred. No. 1.3;
  Matches   99;  Conservative    0;  Mismatches   82;  Indels    0;  Gaps    0;

Qy       8 TTTATGTATTTATCTATAGCATTAGAAACTATAAGAGTTGTTAGCTTCACTTGGCTTACT 67
           ||  | ||||  | |   | |||| |    ||  || |||||  || |  | || | ||
Db   98161 TTATTATATTGGTATGATGTATTACATTAATTATCAGATGTTAAATTAATATTGCATTCT 98220

Qy      68 GTTGTGCTCAAAGCAACTTCATCATCATACAGTATGGTTTTGATATGCTCTTCCATTATC 127
           | || | | ||  | ||||  | ||  | ||  ||  |||| ||||| | ||  ||||
Db   98221 GCTGGGGTAAATCCTACTTGGTTATGGTCCATAATACTTTTTATATGTTGTTGGATTAGG 98280

Qy     128 ACTGAGCCTTATGATTATGTTTTACGAGCTTATAATATCACTGATGGTGATTCAGTATTG 187
             ||       |||  || |||  |   ||  |||||  |  ||  | | |  ||| ||
Db   98281 TTTGTTAGCATTGAGGATTTTTACCTCTCTATTAATAAGAGAGAGTGGGCTCTAGTGTTT 98340

Qy     188 T 188
           |
Db   98341 T 98341


RESULT 15
AC079417/c
LOCUS       AC079417               130540 bp    DNA     linear   HTG 01-SEP-2000
DEFINITION  Mus musculus chromosome 16 clone RP23-102A24, WORKING DRAFT
            SEQUENCE, 34 unordered pieces.
ACCESSION   AC079417
VERSION     AC079417.1  GI:9958029
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Mus musculus.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 130540)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Sequencing of Mouse
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 130540)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-SEP-2000) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT     -------Genome Center
            Center: Joint Genome Institute
            Center Code: JGI
            Web site: http://www.jgi.doe.gov
            -------
            Project Information
            Center Project Name: 0
            Center clone name: RPCI-23_102A24
            -------
            Summary Statistics
            Consensus quality: 71349 bases at least Q40
            Consensus quality: 83788 bases at least Q30
            Consensus quality: 89369 bases at least Q20
            Estimated insert size: 93300; agarose-fp estimation
            Estimated insert size: 127240; sum-of-contigs estimation
            Quality coverage:  3.69 in Q20 bases; agarose-fp estimation
            Quality coverage:  2.7 in Q20 bases; sum-of-contigs estimation.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 34 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1     1043: contig of 1043 bp in length
            *    1044     1143: gap of unknown length
            *    1144     2160: contig of 1017 bp in length
            *    2161     2260: gap of unknown length
            *    2261     3696: contig of 1436 bp in length
            *    3697     3796: gap of unknown length
            *    3797     4826: contig of 1030 bp in length
            *    4827     4926: gap of unknown length
            *    4927     5959: contig of 1033 bp in length
            *    5960     6059: gap of unknown length
            *    6060     7579: contig of 1520 bp in length
            *    7580     7679: gap of unknown length
            *    7680     9178: contig of 1499 bp in length
            *    9179     9278: gap of unknown length
```

```
                *     9279    10501: contig of 1223 bp in length
                *    10502    10601: gap of unknown length
                *    10602    11610: contig of 1009 bp in length
                *    11611    11710: gap of unknown length
                *    11711    12840: contig of 1130 bp in length
                *    12841    12940: gap of unknown length
                *    12941    14262: contig of 1322 bp in length
                *    14263    14362: gap of unknown length
                *    14363    15984: contig of 1622 bp in length
                *    15985    16084: gap of unknown length
                *    16085    17502: contig of 1418 bp in length
                *    17503    17602: gap of unknown length
                *    17603    18902: contig of 1300 bp in length
                *    18903    19002: gap of unknown length
                *    19003    20008: contig of 1006 bp in length
                *    20009    20108: gap of unknown length
                *    20109    21372: contig of 1264 bp in length
                *    21373    21472: gap of unknown length
                *    21473    22675: contig of 1203 bp in length
                *    22676    22775: gap of unknown length
                *    22776    24088: contig of 1313 bp in length
                *    24089    24188: gap of unknown length
                *    24189    25354: contig of 1166 bp in length
                *    25355    25454: gap of unknown length
                *    25455    26641: contig of 1187 bp in length
                *    26642    26741: gap of unknown length
                *    26742    29340: contig of 2599 bp in length
                *    29341    29440: gap of unknown length
                *    29441    31609: contig of 2169 bp in length
                *    31610    31709: gap of unknown length
                *    31710    33695: contig of 1986 bp in length
                *    33696    33795: gap of unknown length
                *    33796    35117: contig of 1322 bp in length
                *    35118    35217: gap of unknown length
                *    35218    38766: contig of 3549 bp in length
                *    38767    38866: gap of unknown length
                *    38867    43795: contig of 4929 bp in length
                *    43796    43895: gap of unknown length
                *    43896    45218: contig of 1323 bp in length
                *    45219    45318: gap of unknown length
                *    45319    50886: contig of 5568 bp in length
                *    50887    50986: gap of unknown length
                *    50987    57066: contig of 6080 bp in length
                *    57067    57166: gap of unknown length
                *    57167    64436: contig of 7270 bp in length
                *    64437    64536: gap of unknown length
                *    64537    74858: contig of 10322 bp in length
                *    74859    74958: gap of unknown length
                *    74959    86302: contig of 11344 bp in length
                *    86303    86402: gap of unknown length
                *    86403    96560: contig of 10158 bp in length
                *    96561    96660: gap of unknown length
                *    96661   130540: contig of 33880 bp in length.
FEATURES             Location/Qualifiers
     source          1. .130540
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /chromosome="16"
                     /clone="RP23-102A24"
                     /clone_lib="RPCI mouse BAC library 23"
BASE COUNT    37619 a  23855 c  30292 g  28886 t   9888 others
ORIGIN

  Query Match             4.1%;  Score 49.8;  DB 2;  Length 130540;
  Best Local Similarity   44.1%;  Pred. No. 1.3;
  Matches  281;  Conservative    0;  Mismatches  350;  Indels    6;  Gaps    2;

Qy       7 ATTTATGTATTTATCTATAGCATTAGAAACTATAAGAGTTGTTAGCTTCACTTGGCTTAC 66
           |||||  | ||  | | |    ||  ||| ||||    || |     ||| |    ||
Db   23411 ATTTAGTTTTTACTTTTTGAATTTTTAAAATATATTTATTATCTTTATCATTATATTTTT 23352

Qy      67 TGTTGTGCTCAAAGCAACTTCATCATCATACAGTATGGTTTTGATATGCTCTTCCATTAT 126
           |  | | |   |     |   | ||  ||  |  ||  ||| || |  | ||   | ||

Db  23351 TTATATATTTCTATTTTTTATTTTATATTATTGATTGTGTTTAATTTAATTTTATTTAAT 23292

Qy     127 CACTGAGCCTTATGATTATGTTTTACGAGCTTATAATATCACTGATGGTGATTCAGTATT 186
           | |  |   | |   ||| || |  |   ||   || |   |  |  | ||    | |
Db  23291 TATTTTGATGTTTATGTATATTATTTGTTATTTATATTTATTTTTTTATTATCTTTTTCT 23232

Qy     187 GTGATTATGTCCTTCGTTGATTATTCTGTTTCATACAAGTCGTGTAATTTGCTGTTTGTG 246
           |   | | |  ||  ||  ||||  | | |  ||     |  | |  ||| |  ||||
Db  23231 TTTTATTTATTTTTTTTTTTTTATGTTTTATTTTAATTTTTTTTTTTTTTTTTATTGTG 23172

Qy     247 ACAGTACGATAGATCGACTCAACCTTCTGAGGTATTAGTTGAAGTTCATG-TAAATTAGC 305
           | ||   |   | | |      || |  | | ||   |   |  || |    ||
Db  23171 TTAATATTTTTTTTATATTTTTTTTTTTATGTTTTTTTATTTTTTATTTGATTTTATATT 23112

Qy     306 TTTGTTTATCATAGTAGCATTTGATTATTGATGCTCTGTAGCTAATGATAAGCCATTGGA 365
           || || || ||  ||  |||| ||||| ||  | | ||  |  |          ||
Db  23111 ATTTTTAATTATTTTATTATTTATTTATTTATTTTTTTTACTTTTTTTATTTTTTTTTTT 23052

Qy     366 GGGAAGCAAGCTTTCTAAATGAATCTACGAATGGATGATAAAGTTCATGAATATTTTTGT 425
           ||        ||| ||  |   |      | |   | | |     |  | ||||||   |
Db  23051 TTGATT-----TTTTTATTTTTTTGATTTATTTTTTTTTTTTTATTTTATATATTTAATT 22997

Qy     426 TACTTCTGCAGTCAGATCATGAGTTATTGAGTCTATTGTTTTTTTAAGCCTGTTTCAGAT 485
           || ||      |    | ||  ||| |   | |  ||| ||||| ||  | |||  |
Db  22996 TATTTTATTTTTTTTTTTATTTTTAGTTTTTTTTATGTATTTTTTAGTTTTTTTATGTA 22937

Qy     486 GATCCATCATCAGTAACAACATACACGGTGTAGTCCCAAATCCATCATATGCACCTTCTT 545
           | ||  |  |  | |    ||  |   | |        |   ||   |  |||| ||
Db  22936 GTTTTTTATTTTTTTATTTAATTTAAATTNNANATTTTTTGGGTCCCCTTTCCCTTTTT 22877

Qy     546 TTCTTCAATTTGGTCTTGTTTTTTTTTTTTCATGATGTCATTGAATTATTCAAGAAGTCA 605
           ||   |||    |||| ||||||||||||  |  |    |||  || || || ||
Db  22876 TTTAAAAATCCACTCTTTTTTTTTTTTTTTTTTTTTNNTTTTGGGTTTTTTAAAAATATG 22817

Qy     606 CTTCGAGCATAATGATTTTTCAAAATCCACCTTTGTT 642
           |        |||  ||||| | ||  |   ||  ||
Db  22816 TTNTTTTTTAAATATTTTTTTATAAANCCATTTNTTT 22780


Search completed: May 24, 2003, 07:52:07
Job time : 4225 secs
```

GenCore version 5.1.5

OM nucleic - nucleic search, using sw model

Run on: May 24, 2003, 06:35:14 ; Search time 84 Seconds
(without alignments)
4443.162 Million cell updates/sec

Title: US-09-701-926B-1
Perfect score: 1217
Sequence: 1 tttgaaatttatgtatttat.....gtcaacacaacacaacaaca 1217

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	46.4	3.8	7218	1	US-08-232-463-14	Sequence 14, Appl
	2	41.6	3.4	51952	3	US-08-947-823-1	Sequence 1, Appli
	3	40.8	3.4	5718	3	US-08-714-918-48	Sequence 48, Appl
	4	40.8	3.4	5718	4	US-09-265-315-48	Sequence 48, Appl
	5	40.8	3.4	5718	4	US-09-265-315-48	Sequence 48, Appl
	6	40.8	3.4	5718	4	US-09-266-417-48	Sequence 48, Appl
c	7	39.6	3.3	2093	4	US-09-063-733A-42	Sequence 42, Appl
c	8	39.2	3.2	19124	2	US-08-487-826B-13	Sequence 13, Appl
c	9	39	3.2	2610	4	US-09-545-814-1	Sequence 1, Appli
	10	39	3.2	2610	4	US-09-545-814-3	Sequence 3, Appli
	11	39	3.2	5852	1	US-07-867-106-2	Sequence 2, Appli
	12	38.2	3.1	1328	2	US-08-766-439-39	Sequence 39, Appl
c	13	38.2	3.1	1328	2	US-08-766-439-40	Sequence 40, Appl
c	14	38.2	3.1	1813	4	US-09-071-224-3	Sequence 3, Appli
c	15	38	3.1	3095	6	5231168-1	Patent No. 5231168
c	16	37.8	3.1	1334	3	US-08-884-324-8	Sequence 8, Appli
c	17	37.8	3.1	11464	3	US-08-884-324-13	Sequence 13, Appl
c	18	37.8	3.1	28994	3	US-08-884-324-14	Sequence 14, Appl
	19	36.8	3.0	2843	4	US-09-463-238-1	Sequence 1, Appli
	20	36.8	3.0	2843	4	US-09-463-238-10	Sequence 10, Appl
	21	36	3.0	162450	4	US-09-345-882-1	Sequence 1, Appli
c	22	35.6	2.9	987	4	US-09-142-584-1	Sequence 1, Appli
c	23	35.6	2.9	987	4	US-09-142-584-3	Sequence 3, Appli
c	24	35.6	2.9	987	4	US-09-142-584-5	Sequence 5, Appli
c	25	35.6	2.9	4140	3	US-08-894-731-2	Sequence 2, Appli
c	26	35.4	2.9	834	4	US-08-998-416-347	Sequence 347, App
c	27	35.4	2.9	863	4	US-08-998-416-548	Sequence 548, App
c	28	35.4	2.9	9048	3	US-08-973-273-4	Sequence 4, Appli
	29	35	2.9	470	4	US-09-020-956-102	Sequence 102, App
	30	35	2.9	470	4	US-09-030-607-102	Sequence 102, App
	31	35	2.9	470	4	US-09-605-785-102	Sequence 102, App
	32	35	2.9	470	4	US-09-439-313-102	Sequence 102, App
	33	35	2.9	470	4	US-09-352-616A-102	Sequence 102, App
	34	35	2.9	470	4	US-09-232-149A-102	Sequence 102, App
	35	35	2.9	2882	1	US-08-393-219-2	Sequence 2, Appli
c	36	35	2.9	30549	4	US-09-134-001C-322	Sequence 322, App
c	37	34.8	2.9	748	1	US-08-361-467B-3	Sequence 3, Appli
c	38	34.8	2.9	748	1	US-08-484-332C-3	Sequence 3, Appli
c	39	34.8	2.9	3483	4	US-09-130-491-3	Sequence 3, Appli
c	40	34.8	2.9	36651	4	US-09-738-894A-3	Sequence 3, Appli
c	41	34.6	2.8	236	4	US-09-889-595-8	Sequence 8, Appli
c	42	34.6	2.8	823	4	US-09-097-889-3	Sequence 3, Appli
c	43	34.6	2.8	1263	4	US-09-302-681-4	Sequence 4, Appli
c	44	34.6	2.8	5058	4	US-09-889-595-1	Sequence 1, Appli
c	45	34.2	2.8	168575	4	US-09-426-290-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-08-232-463-14
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTZgpt-F1s
US-08-232-463-14

Query Match 3.8%; Score 46.4; DB 1; Length 7218;

```
  Best Local Similarity   7.3%;  Pred. No. 0.021;
  Matches   32;  Conservative  214;  Mismatches  190;  Indels    0;  Gaps    0;

Qy     389 TCTACGAATGGATGATAAAGTTCATGAATATTTTTGTTACTTCTGCAGTCAGATCATGAG 448
           | | | |   | | | | | || ||    |  || | | |   |  ||  || | |   ||
Db    1007 TTTGCCATACGCTCACAGAATTAATTCCGAGCTTGGCTGCAGGTCGAGGGAGCTTGCGAT 1066

Qy     449 TTATTGAGTCTATTGTTTTTTTAAGCCTGTTTCAGATGATCCATCATCAGTAACAACATA 508
           :: ::   ::: :: :::::::   ::: ::::   :  ::: :: ::  :  :  : :
Db    1067 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1126

Qy     509 CACGGTGTAGTCCCAAATCCATCATATGCACCTTCTTTTCTTCAATTTGGTCTTGTTTTT 568
           : :  : :  ::::   ::: :: : : : :::::::::::::  :::  :::: :::::
Db    1127 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1186

Qy     569 TTTTTTTCATGATGTCATTGAATTATTCAAGAAGTCACTTCGAGCATAATGATTTTTCAA 628
           :::::::: :  : :: ::   :: :::      :: ::::   : :  :  ::::::
Db    1187 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1246

Qy     629 AATCCACCTTTGTTCAAGCACTACCACGTCTTTTCATCTAGCCCACAACCGTGGTGGAGG 688
           :::: ::::: :::   : :: :: : :::::::: :::  ::: :  :: :  :
Db    1247 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1306

Qy     689 ATCTAGAATTTTCATGAAAGGATTCAAAATTTACAAACATATATATACACTATACACTAT 748
           :::    ::::: :       :::   ::: :   : : : : : : :: : : :: :
Db    1307 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1366

Qy     749 GAATCCACTAATACTAGATGGTGCACCTGTGCCCCCACTCATGTGAAAGCCTATTCTCAA 808
           ::: ::  :  ::   :  : : ::: : ::::: ::: : :     ::: :::::
Db    1367 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1426

Qy     809 TTTTTTATTTTCCACA 824
           :::::: : | ||| |
Db    1427 YYYYYYYYGTACCAAA 1442


RESULT 2
US-08-947-823-1
; Sequence 1, Application US/08947823
; Patent No. 6114605
;  GENERAL INFORMATION:
;    APPLICANT:  Williamson, Valerie M.
;    APPLICANT:  Kaloshian, Isgouhi
;    APPLICANT:  Yaghoobi, Jafar
;    APPLICANT:  Bodeau, John
;    APPLICANT:  Milligan, Stephen
;    TITLE OF INVENTION:  Procedures and Materials for Conferring
;    TITLE OF INVENTION:  Pest Resistance in Plants
;    NUMBER OF SEQUENCES:  5
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Townsend and Townsend and Crew LLP
;      STREET:  Two Embarcadero Center, Eighth Floor
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94111-3834
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/947,823
;      FILING DATE:  09-OCT-1997
;      CLASSIFICATION:  800
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US97/18802
;      FILING DATE:  09-OCT-1997
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/028,191
;      FILING DATE:  10-OCT-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Bastian, Kevin L.
;      REGISTRATION NUMBER:  34,774
;      REFERENCE/DOCKET NUMBER:  023070-070210US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 576-0200
;      TELEFAX:  (415) 576-0300
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  51952 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-947-823-1

  Query Match             3.4%;  Score 41.6;  DB 3;  Length 51952;
  Best Local Similarity   47.1%;  Pred. No. 0.63;
  Matches  160;  Conservative    0;  Mismatches  179;  Indels    1;  Gaps    1;

Qy     686 AGGATCTAGAATTTTCATGAAAGGATTCAAAATTTACAAACATATATATACACTATACAC 745
           || |  |   ||||| || ||    ||||| ||||| ||    | ||    |  ||
Db   33273 AGAAATTGATATTTTAATAAATATTTTCAATATTTAAAATATAACATTATGATAATTTGT 33332

Qy     746 TA-TGAATCCACTAATACTAGATGGTGCACCTGTGCCCCCACTCATGTGAAAGCCTATTC 804
           || || || |  ||||      | ||    ||     | |  |||| |||   |
Db   33333 TATTGTATACTGAAATAAATTTGATTTCATGATTGTATTTAGTGTTGTGGAAGTTAAAAA 33392

Qy     805 TCAATTTTTTATTTTCCACAACTTAAATACAGACCGCACAACTCCCGTGTCTTGTGTGCT 864
           |||| ||||||  |   ||     | ||  | |||  |  |        || || | |  |
Db   33393 TCAAGTTTTTAATAAATACTTTCAATATTAATACCAAAATAAATAAAAGTATTCTATTAT 33452

Qy     865 CGTCGCTCAGCATGCAAGTCGAGAAAAGAAAGACCAAAACAATGAAAACTTTACGAAAAA 924
           ||  |   |  ||| ||      |||| |   ||||| | |   ||      ||  | |
Db   33453 CGAAGACAATAATGAAATGTATCAAAAAATTCACCAACAAAGCAAACTACAAACACATTA 33512

Qy     925 TCAAAAAGTTGAAGGACTTTAACGTCGAGATCTCTCGTAGAAAACCTCTTTTGTAAGGTT 984
           |    |||||||||  | |  |   |  | | |   | | |   | | |   ||   | |
Db   33513 TACTTAAGTTGAAGCAATAGAGATTTTAAAGGCTTTGAATCGATCTTTACGTGGTTGATC 33572

Qy     985 GCATACAATACTTTTTTTTCAGACTTTACTTATGGTATTA 1024
             |  |   |||| |   | ||    ||||||||||
Db   33573 TATGATTAGTATTTTATCAGATACAAAGCTTATGGTATAA 33612


RESULT 3
US-08-714-918-48
; Sequence 48, Application US/08714918
; Patent No. 6037123
;  GENERAL INFORMATION:
;    APPLICANT:  Benton, Bret
;    APPLICANT:  Lee, Ving
;    APPLICANT:  Malouin, Francois
;    APPLICANT:  Martin, Patrick K.
;    APPLICANT:  Schmid, Molly B.
;    APPLICANT:  Sun, Dongxu
;    TITLE OF INVENTION:  STAPHYLOCOCCUS AUREUS ANTIBACTERIAL
;    TITLE OF INVENTION:  TARGET GENES
;    NUMBER OF SEQUENCES:  111
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Lyon & Lyon
;      STREET:  633 West Fifth Street
;      STREET:  Suite 4700
;      CITY:  Los Angeles
;      STATE:  California
;      COUNTRY:  U.S.A.
;      ZIP:  90071-2066
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5" Diskette, 1.44 Mb
;      MEDIUM TYPE:  storage
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  IBM P.C. DOS 5.0
;      SOFTWARE:  Word Perfect 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/714,918
```

```
;      FILING DATE:  September 13, 1996
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  60/009,102
;      FILING DATE:  December 22, 1995
;      APPLICATION NUMBER:  60/003,798
;      FILING DATE:  September 15, 1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Warburg, Richard J.
;      REGISTRATION NUMBER:  32,327
;      REFERENCE/DOCKET NUMBER:  222/005
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (213) 489-1600
;      TELEFAX:  (213) 955-0440
;      TELEX:  67-3510
;  INFORMATION FOR SEQ ID NO:  48:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  5718 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-714-918-48

  Query Match             3.4%;  Score 40.8;  DB 3;  Length 5718;
  Best Local Similarity   53.0%;  Pred. No. 0.48;
  Matches   87;  Conservative    0;  Mismatches   77;  Indels    0;  Gaps    0;

Qy     381 TAAATGAATCTACGAATGGATGATAAAGTTCATGAATATTTTTGTTACTTCTGCAGTCAG 440
           ||| || |||    |||   |  | |  |||||  || | |||     ||       |||
Db    1532 TAAGTGCATCAGGCAATTCTTCTTTATCTTCATACATCTGTTTTGGCGTTAAAATATCAA 1591

Qy     441 ATCATGAGTTATTGAGTCTATTGTTTTTTTAAGCCTGTTTCAGATGATCCATCATCAGTA 500
           || |  |   ||   ||  | |||||   |   | |    ||  ||||| ||     | |
Db    1592 ATAAAAAACCATCTTGTTCAATGTTTAAATCGCCATCCAACACTTGATCAATAGCTTGCA 1651

Qy     501 ACAACATACACGGTGTAGTCCCAAATCCATCATATGCACCTTCT 544
           |||||  |  |  ||||  | |||||  ||||| | || ||  ||
Db    1652 ACAACGTTGAACGTGTTTTACCAAAAGCATCAAACGCTCCCACT 1695


RESULT 4
US-09-265-315-48
; Sequence 48, Application US/09265315
; Patent No. 6187541
;  GENERAL INFORMATION:
;    APPLICANT:  Benton, Bret
;    APPLICANT:  Lee, Ving J.
;    APPLICANT:  Malouin, Francois
;    APPLICANT:  Martin, Patrick K.
;    APPLICANT:  Schmid, Molly B.
;    APPLICANT:  Sun, Dongxu
;    TITLE OF INVENTION:  METHODS OF SCREENING FOR COMPOUNDS
;    TITLE OF INVENTION:  ACTIVE ON STAPHYLOCOCCUS AUREUS
;    TITLE OF INVENTION:  TARGET GENES
;    NUMBER OF SEQUENCES:  111
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Lyon & Lyon
;      STREET:  633 West Fifth Street
;      STREET:  Suite 4700
;      CITY:  Los Angeles
;      STATE:  California
;      COUNTRY:  U.S.A.
;      ZIP:  90071-2066
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5" Diskette, 1.44 Mb
;      MEDIUM TYPE:  storage
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  IBM P.C. DOS 5.0
;      SOFTWARE:  Word Perfect 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/265,315
;      FILING DATE:  March 9, 1999
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/714,918
;      FILING DATE:  September 13, 1996
;      APPLICATION NUMBER:  60/009,102
;      FILING DATE:  December 22, 1995
;      APPLICATION NUMBER:  60/003,798
;      FILING DATE:  September 15, 1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Warburg, Richard J.
;      REGISTRATION NUMBER:  32,327
;      REFERENCE/DOCKET NUMBER:  240/247
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (213) 489-1600
;      TELEFAX:  (213) 955-0440
;      TELEX:  67-3510
;  INFORMATION FOR SEQ ID NO:  48:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  5718 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-09-265-315-48

  Query Match             3.4%;  Score 40.8;  DB 4;  Length 5718;
  Best Local Similarity   53.0%;  Pred. No. 0.48;
  Matches   87;  Conservative    0;  Mismatches   77;  Indels    0;  Gaps    0;

Qy     381 TAAATGAATCTACGAATGGATGATAAAGTTCATGAATATTTTTGTTACTTCTGCAGTCAG 440
           ||| || |||    |||   |  | |  |||||  || | |||     ||       |||
Db    1532 TAAGTGCATCAGGCAATTCTTCTTTATCTTCATACATCTGTTTTGGCGTTAAAATATCAA 1591

Qy     441 ATCATGAGTTATTGAGTCTATTGTTTTTTTAAGCCTGTTTCAGATGATCCATCATCAGTA 500
           || |  |   ||   ||  | |||||   |   | |    ||  ||||| ||     | |
Db    1592 ATAAAAAACCATCTTGTTCAATGTTTAAATCGCCATCCAACACTTGATCAATAGCTTGCA 1651

Qy     501 ACAACATACACGGTGTAGTCCCAAATCCATCATATGCACCTTCT 544
           |||||  |  |  ||||  | |||||  ||||| | || ||  ||
Db    1652 ACAACGTTGAACGTGTTTTACCAAAAGCATCAAACGCTCCCACT 1695


RESULT 5
US-09-265-315-48
; Sequence 48, Application US/09265315
; Patent No. 6187541
;  GENERAL INFORMATION:
;    APPLICANT:  Benton, Bret
;    APPLICANT:  Lee, Ving J.
;    APPLICANT:  Malouin, Francois
;    APPLICANT:  Martin, Patrick K.
;    APPLICANT:  Schmid, Molly B.
;    APPLICANT:  Sun, Dongxu
;    TITLE OF INVENTION:  METHODS OF SCREENING FOR COMPOUNDS
;    TITLE OF INVENTION:  ACTIVE ON STAPHYLOCOCCUS AUREUS
;    TITLE OF INVENTION:  TARGET GENES
;    NUMBER OF SEQUENCES:  111
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Lyon & Lyon
;      STREET:  633 West Fifth Street
;      STREET:  Suite 4700
;      CITY:  Los Angeles
;      STATE:  California
;      COUNTRY:  U.S.A.
;      ZIP:  90071-2066
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5" Diskette, 1.44 Mb
;      MEDIUM TYPE:  storage
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  IBM P.C. DOS 5.0
;      SOFTWARE:  Word Perfect 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/265,315
;      FILING DATE:  March 9, 1999
;      CLASSIFICATION:  435
```

```
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/714,918
;      FILING DATE:  September 13, 1996
;      APPLICATION NUMBER:  60/009,102
;      FILING DATE:  December 22, 1995
;      APPLICATION NUMBER:  60/003,798
;      FILING DATE:  September 15, 1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Warburg, Richard J.
;      REGISTRATION NUMBER:  32,327
;      REFERENCE/DOCKET NUMBER:  240/247
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (213) 489-1600
;      TELEFAX:  (213) 955-0440
;      TELEX:  67-3510
;  INFORMATION FOR SEQ ID NO:  48:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  5718 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-09-265-315-48

  Query Match             3.4%;  Score 40.8;  DB 4;  Length 5718;
  Best Local Similarity   53.0%;  Pred. No. 0.48;
  Matches   87;  Conservative    0;  Mismatches   77;  Indels    0;  Gaps    0;

Qy     381 TAAATGAATCTACGAATGGATGATAAAGTTCATGAATATTTTTGTTACTTCTGCAGTCAG 440
           ||| || |||    |||   |  | |  ||||  || | |||     ||      |||
Db    1532 TAAGTGCATCAGGCAATTCTTCTTTATCTTCATACATCTGTTTTGGCGTTAAAATATCAA 1591

Qy     441 ATCATGAGTTATTGAGTCTATTGTTTTTTTAAGCCTGTTTCAGATGATCCATCATCAGTA 500
           || |  |   ||   ||  | |||||   |   | |    ||  |||||  ||     | |
Db    1592 ATAAAAAACCATCTTGTTCAATGTTTAAATCGCCATCCAACACTTGATCAATAGCTTGCA 1651

Qy     501 ACAACATACACGGTGTAGTCCCAAATCCATCATATGCACCTTCT 544
           |||||  |  |  ||||  | |||||  |||||  | || ||  ||
Db    1652 ACAACGTTGAACGTGTTTTACCAAAAGCATCAAACGCTCCCACT 1695


RESULT 6
US-09-266-417-48
; Sequence 48, Application US/09266417
; Patent No. 6228588
;  GENERAL INFORMATION:
;    APPLICANT:  Benton, Bret
;    APPLICANT:  Lee, Ving J.
;    APPLICANT:  Malouin, Francois
;    APPLICANT:  Martin, Patrick K.
;    APPLICANT:  Schmid, Molly B.
;    APPLICANT:  Sun, Dongxu
;    TITLE OF INVENTION:  METHODS OF SCREENING FOR COMPOUNDS
;    TITLE OF INVENTION:  ACTIVE ON STAPHYLOCOCCUS AUREUS
;    TITLE OF INVENTION:  TARGET GENES
;    NUMBER OF SEQUENCES:  111
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Lyon & Lyon
;      STREET:  633 West Fifth Street
;      STREET:  Suite 4700
;      CITY:  Los Angeles
;      STATE:  California
;      COUNTRY:  U.S.A.
;      ZIP:  90071-2066
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5" Diskette, 1.44 Mb
;      MEDIUM TYPE:  storage
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  IBM P.C. DOS 5.0
;      SOFTWARE:  Word Perfect 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/266,417
;      FILING DATE:  March 9, 1999
;      CLASSIFICATION:  435
```

```
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/714,918
;      FILING DATE:  September 13, 1996
;      APPLICATION NUMBER:  60/009,102
;      FILING DATE:  December 22, 1995
;      APPLICATION NUMBER:  60/003,798
;      FILING DATE:  September 15, 1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Warburg, Richard J.
;      REGISTRATION NUMBER:  32,327
;      REFERENCE/DOCKET NUMBER:  240/248
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (213) 489-1600
;      TELEFAX:  (213) 955-0440
;      TELEX:  67-3510
;  INFORMATION FOR SEQ ID NO:  48:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  5718 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-09-266-417-48

  Query Match             3.4%;  Score 40.8;  DB 4;  Length 5718;
  Best Local Similarity   53.0%;  Pred. No. 0.48;
  Matches   87;  Conservative    0;  Mismatches   77;  Indels    0;  Gaps    0;

Qy     381 TAAATGAATCTACGAATGGATGATAAAGTTCATGAATATTTTTGTTACTTCTGCAGTCAG 440
           ||| || |||    |||   |  | |  ||||  || | |||     ||      |||
Db    1532 TAAGTGCATCAGGCAATTCTTCTTTATCTTCATACATCTGTTTTGGCGTTAAAATATCAA 1591

Qy     441 ATCATGAGTTATTGAGTCTATTGTTTTTTTAAGCCTGTTTCAGATGATCCATCATCAGTA 500
           || |  |   ||   ||  | |||||   |   | |    ||  |||||  ||     | |
Db    1592 ATAAAAAACCATCTTGTTCAATGTTTAAATCGCCATCCAACACTTGATCAATAGCTTGCA 1651

Qy     501 ACAACATACACGGTGTAGTCCCAAATCCATCATATGCACCTTCT 544
           |||||  |  |  ||||  | |||||  |||||  | || ||  ||
Db    1652 ACAACGTTGAACGTGTTTTACCAAAAGCATCAAACGCTCCCACT 1695


RESULT 7
US-09-063-733A-42/c
; Sequence 42, Application US/09063733A
; Patent No. 6372211
;  GENERAL INFORMATION:
;    APPLICANT:  Isaac, Barbara G.
;    APPLICANT:  Greenplate, John T.
;    APPLICANT:  Purcell, John P.
;    APPLICANT:  Romano, Charles P.
;    TITLE OF INVENTION:  METHODS AND COMPOSITIONS FOR CONTROLLING
;    TITLE OF INVENTION:  INSECTS
;    NUMBER OF SEQUENCES:  58
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Arnold White & Durkee
;      STREET:  PO Box 4433
;      CITY:  Houston
;      STATE:  TX
;      COUNTRY:  USA
;      ZIP:  77210-4433
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/063,733A
;      FILING DATE:  21-APR-1998
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Patterson, Melinda L.
;      REGISTRATION NUMBER:  33,062
;      REFERENCE/DOCKET NUMBER:  MOBT:022
;    TELECOMMUNICATION INFORMATION:
```

```
;       TELEPHONE:  713-787-1400
;       TELEFAX:  713-787-1440
;   INFORMATION FOR SEQ ID NO:  42:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  2093 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
US-09-063-733A-42

  Query Match             3.3%;  Score 39.6;  DB 4;  Length 2093;
  Best Local Similarity   51.1%;  Pred. No. 0.68;
  Matches   93;  Conservative    0;  Mismatches   89;  Indels    0;  Gaps    0;

Qy      13 GTATTTATCTATAGCATTAGAAACTATAAGAGTTGTTAGCTTCACTTGGCTTACTGTTGT 72
           | ||| || ||   ||| |     | |     |  |||       |||    ||||| |
Db    2077 GGATTCATATACCTCATCAATTCTTTTTCACCTATTTATGCATCTTTGCAACACTGTAGA 2018

Qy      73 GCTCAAAGCAACTTCATCATCATACAGTATGGTTTTGATATGCTCTTCCATTATCACTGA 132
             |  | ||||||  |  ||   || |||||| ||| |||  |    |||   |
Db    2017 CCAGATAGCAACACCTACAAGCTATAGTATGCTTTGCGTATTCTCAACAGCGATCTTGGT 1958

Qy     133 GCCTTATGATTATGTTTTACGAGCTTATAATATCACTGATGGTGATTCAGTATTGTGATT 192
             ||| | ||| ||| || |     |||  |||| |   || | |  | ||||   ||
Db    1957 CACTTCTCATTGTGTGTTCCACAGATATCCTATCCCGCCTGCTTACCCCCTATTCACTTT 1898

Qy     193 AT 194
           |
Db    1897 CT 1896


RESULT 8
US-08-487-826B-13/c
; Sequence 13, Application US/08487826B
; Patent No. 5993827
;  GENERAL INFORMATION:
;     APPLICANT:  Sim, Kim L.
;     APPLICANT:  Chitnis, Chetan
;     APPLICANT:  Miller, Louis H.
;     APPLICANT:  Peterson, David S.
;     APPLICANT:  Su, Xin-zhaun
;     APPLICANT:  Wellems, Thomas E.
;     TITLE OF INVENTION:  BINDING DOMAINS FROM PLASMODIUM VIVAX
;     TITLE OF INVENTION:  AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
;     NUMBER OF SEQUENCES:  45
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Knobbe Martens Olson & Bear
;       STREET:  620 Newport Center Drive  16th Floor
;       CITY:  Newport Beach
;       STATE:  California
;       COUNTRY:  US
;       ZIP:  92660
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/487,826B
;       FILING DATE:  10-SEP-1993
;       CLASSIFICATION:  435
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Israelsen, Ned
;       REGISTRATION NUMBER:  29,655
;       REFERENCE/DOCKET NUMBER:  NIH121.001CP1
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (619) 235-8550
;       TELEFAX:  (619) 235-0176
;   INFORMATION FOR SEQ ID NO:  13:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  19124 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
;     HYPOTHETICAL:  NO
;     ANTI-SENSE:  NO
US-08-487-826B-13

  Query Match             3.2%;  Score 39.2;  DB 2;  Length 19124;
  Best Local Similarity   47.2%;  Pred. No. 1.8;
  Matches  119;  Conservative    0;  Mismatches  133;  Indels    0;  Gaps    0;

Qy     909 AAAACTTTACGAAAAATCAAAAAGTTGAAGGACTTTAACGTCGAGATCTGTCGTAGAAAA 968
           || | | | |  | |||||| |  ||  || ||  |||  | |  |   |     ||||
Db    4436 AATAATATTCTCATAATCAATATATTTTAGAACAATAAAATAAAATTAAATTAATAAAAA 4377

Qy     969 CCTCTTTTGTAAGGTTGCATACAATACTTTTTTTTCAGACTTTACTTATGGTATTATACT 1028
             |||  |  |            || |||||| | | | | |   ||   | | || |
Db    4376 AAAATTTATTGTGACATTTATGCTTAATTTTTTATAATATTATTTATAATATTTCTTATT 4317

Qy    1029 GAATATGTTATTGCTGTTATAGTAGTTGAGTGACGTTTGAGGGAATTTCTAGTCCGTTAA 1088
             |||| || ||    ||| | |  |  | | || ||      || |||| |   |  |
Db    4316 ATATATATTTTTTTAATTACAATTATGCAAAGTTGTATGTTAAGATATCTATTATATATA 4257

Qy    1089 TCTTGTACTCAGTGTGTCTACTTTTCAAAAAAGTCAGTTTTTCAGTCTCTAAAACACATT 1148
           ||||  |  || |    |  | | ||||||| |  |          ||  |||||||
Db    4256 TCTTACAAACATTTATAATCAATATTAAAAAAGCCTTTAACAACTATCCTTTTACACATT 4197

Qy    1149 TAAATAAGAGTT 1160
             ||   | ||
Db    4196 CTTATGTTATTT 4185


RESULT 9
US-09-545-814-1/c
; Sequence 1, Application US/09545814
; Patent No. 6416977
; GENERAL INFORMATION:
;  APPLICANT:  Becher, Anna M.
;  TITLE OF INVENTION:  FLEA CHITINASE NUCLEIC ACID MOLECULES, PROTEINS AND
;  TITLE OF INVENTION:  USES THEREOF
;  FILE REFERENCE:  FC-5-C1
;  CURRENT APPLICATION NUMBER:  US/09/545,814
;  CURRENT FILING DATE:  2000-04-07
;  PRIOR APPLICATION NUMBER:  60/128,833
;  PRIOR FILING DATE:  1999-04-09
;  NUMBER OF SEQ ID NOS:  40
;  SOFTWARE:  PatentIn Ver. 2.1
; SEQ ID NO 1
;   LENGTH: 2610
;   TYPE: DNA
;   ORGANISM: Ctenocephalides felis
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)..(1749)
US-09-545-814-1

  Query Match             3.2%;  Score 39;  DB 4;  Length 2610;
  Best Local Similarity   48.4%;  Pred. No. 1;
  Matches  108;  Conservative    0;  Mismatches  115;  Indels    0;  Gaps    0;

Qy     541 TTCTTTTCTTCAATTTGGTCTTGTTTTTTTTTTTTTCATGATGTCATTGAATTATTCAAGA 600
           || |||| ||   |||  | || ||||||||||||  |  | |  ||   || |    |
Db    2607 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTCGTTTTTATTCTA 2548

Qy     601 AGTCACTTCGAGCATAATGATTTTTCAAAATCCACCTTTGTTCAAGCACTACCACGTCTT 660
             | | |   |  ||||  | ||   ||      |  | |  ||| | || |   | | ||
Db    2547 TTTTATTATTAATATAAATAATTACAAATCCAAAATTCTTATCACGTACAATTTCATATT 2488

Qy     661 TTCATCTAGCCCACAACCGTGGTGGAGGATCTAGAATTTTCATGAAAGGATTCAAAATTT 720
           ||  |  | |    |  | ||| |||     | |        |  |  |  | || |
Db    2487 TTTCTACAACGTTTATAAATATTTGAGTATCATATTTGTAGGACCTAATAAACTTAAATA 2428

Qy     721 ACAAACATATATATACACTATACACTATGAATCCACTAATACT 763
           |  ||   |  | | | | || |||  | ||  |||| ||||
```

```
Db    2427 AATAAACAAACAAAAAAATTTATACTTAGGATATACTAGTACT 2385


RESULT 10
US-09-545-814-3
; Sequence 3, Application US/09545814
; Patent No. 6416977
; GENERAL INFORMATION:
;  APPLICANT:  Becher, Anna M.
;  TITLE OF INVENTION:  FLEA CHITINASE NUCLEIC ACID MOLECULES, PROTEINS AND
;  TITLE OF INVENTION:  USES THEREOF
;  FILE REFERENCE:  FC-5-C1
;  CURRENT APPLICATION NUMBER: US/09/545,814
;  CURRENT FILING DATE:  2000-04-07
;  PRIOR APPLICATION NUMBER:  60/128,833
;  PRIOR FILING DATE:  1999-04-09
;  NUMBER OF SEQ ID NOS:  40
;  SOFTWARE:  PatentIn Ver. 2.1
; SEQ ID NO 3
;   LENGTH: 2610
;   TYPE: DNA
;   ORGANISM: Ctenocephalides felis
US-09-545-814-3

  Query Match             3.2%;  Score 39;  DB 4;  Length 2610;
  Best Local Similarity   48.4%;  Pred. No. 1;
  Matches  108;  Conservative    0;  Mismatches  115;  Indels    0;  Gaps    0;

Qy     541 TTCTTTTCTTCAATTTGGTCTTGTTTTTTTTTTTTCATGATGTCATTGAATTATTCAAGA 600
           || |||| ||   |||  | || ||||||||||||  |  | |  ||   || |     |
Db       4 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTCGTTTTATTCTA 63

Qy     601 AGTCACTTCGAGCATAATGATTTTTCAAAATCCACCTTTGTTCAAGCACTACCACGTCTT 660
           | | |   |  ||||  | ||   ||     |  | |  ||| | || |   | | ||
Db      64 TTTTATTATTAATATAAATAATTACAAATCCAAAATTCTTATCACGTACAATTTCATATT 123

Qy     661 TTCATCTAGCCCACAACCGTGGTGGAGGATCTAGAATTTTCATGAAAGGATTCAAAATTT 720
           ||  |  | |    |   | ||| |||    | |       |  |  |  || |
Db     124 TTTCTACAACGTTTATAAATATTTGAGTATCATATTTGTAGGACCTAATAAACTTAAATA 183

Qy     721 ACAAACATATATATACACTATACACTATGAATCCACTAATACT 763
           |  ||   |  | | | | || |||  | ||  |||| ||||
Db     184 AATAAACAAACAAAAAAATTTATACTTAGGATATACTAGTACT 226


RESULT 11
US-07-867-106-2
; Sequence 2, Application US/07867106
; Patent No. 5389526
;  GENERAL INFORMATION:
;    APPLICANT:  Slade, Martin B
;    APPLICANT:  Chang, Andy C M
;    APPLICANT:  Williams, Keith L
;    TITLE OF INVENTION:  Improved Plasmid Vectors for Cellular
;    TITLE OF INVENTION:  Slime Moulds of the Genus Dictyostelium
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Woodcock Washburn Kurtz Mackiewicz & No. 5389526ris
;      STREET:  One Liberty Place 46th Floor
;      CITY:  Philadelphia
;      STATE:  PA
;      COUNTRY:  USA
;      ZIP:  19103
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/867,106
;      FILING DATE:  19920625
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  AU PJ 7187
;      APPLICATION NUMBER:  PCT/AU90/00530
;      FILING DATE:  02-NOV-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Feeney, Joanne Longo
;      REGISTRATION NUMBER:  35,134
;      REFERENCE/DOCKET NUMBER:  RICE-0002
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  215-568-3100
;      TELEFAX:  215-568-3439
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  5852 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    ANTI-SENSE:  NO
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  2378..5038
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  2378..5038
US-07-867-106-2

  Query Match             3.2%;  Score 39;  DB 1;  Length 5852;
  Best Local Similarity   47.1%;  Pred. No. 1.4;
  Matches  120;  Conservative    0;  Mismatches  135;  Indels    0;  Gaps    0;

Qy     900 AAAACAATGAAAACTTTACGAAAAATCAAAAAGTTGAAGGACTTTAACGTCGAGATCTCT 959
           |||| ||  ||||    |  |||||||||| |  |  |    |||||  |      |||
Db    5450 AAAAAAAAAAAAAAAAAAAAAAAAAATCAAATATGTTTATTGTTTTAAGATTTTTAATCT 5509

Qy     960 CGTAGAAAACCTCTTTTGTAAGGTTGCATACAATACTTTTTTTTCAGACTTTACTTATGG 1019
           |||  |  | |       ||  | |      |  ||       |  ||||| | |
Db    5510 CGTCAATGATTTTAAAATAAAAATCGATACATAATTTTAAAAAAAACCCTTTACATTTTT 5569

Qy    1020 TATTATACTGAATATGTTATTGCTGTTATAGTAGTTGAGTGACGTTTGAGGGAATTTCTA 1079
           ||||  || |    |  ||||   | || || |  ||    |  |||       ||| ||
Db    5570 TATTTTAATTCCAAATTTATACATTTTTTATTTTTTTTTTTTTTTTTTTTTTTTTTTTA 5629

Qy    1080 GTCCGTTAATCTTGTACTCAGTGTGTCTACTTTTCAAAAAAGTCAGTTTTTCAGTCTCTA 1139
           |     | ||  |  |   | | | |  |||    |||| | | || || | | ||
Db    5630 ATTTAAATTTTTTTTTTTTTTTTTTTTTTTATTTATTTAAAATTTAATTATTAATTTTATA 5689

Qy    1140 AAACACATTTAAATA 1154
           ||  | |  || | |
Db    5690 AATAAAAAATAGAAA 5704


RESULT 12
US-08-766-439-39
; Sequence 39, Application US/08766439
; Patent No. 5922538
;  GENERAL INFORMATION:
;    APPLICANT:  HAZEL, JAMES WILLIAM
;    APPLICANT:  JENSEN, MARK ANTON
;    TITLE OF INVENTION:  GENETIC MARKERS AND METHODS FOR
;    TITLE OF INVENTION:  THE DETECTION OF LISTERIA
;    TITLE OF INVENTION:  MONOCYTOGENES AND LISTERIA SPP.
;    NUMBER OF SEQUENCES:  110
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  E. I. DU PONT DE NEMOURS AND COMPANY
;      STREET:  1007 MARKET STREET
;      CITY:  WILMINGTON
;      STATE:  DELAWARE
;      COUNTRY:  U.S.A.
;      ZIP:  19898
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.50 INCH DISKETTE
;      COMPUTER:  IBM PC COMPATIBLE
;      OPERATING SYSTEM:  MICROSOFT WINDOWS 3.1
;      SOFTWARE:  MICROSOFT WORD 2.0C
```

```
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/766,439
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/745,228
;      FILING DATE:  NOVEMBER 8, 1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  FLOYD, LINDA AXAMETHY
;      REGISTRATION NUMBER:  33,692
;      REFERENCE/DOCKET NUMBER:  MD-1065-A
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  302-892-8112
;      TELEFAX:  302-773-0164
;  INFORMATION FOR SEQ ID NO:  39:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1328 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    ORIGINAL SOURCE:
;      STRAIN:  L IVANOVII 3340 D.F.
US-08-766-439-39

  Query Match             3.1%;  Score 38.2;  DB 2;  Length 1328;
  Best Local Similarity   54.8%;  Pred. No. 1.3;
  Matches  121;  Conservative    0;  Mismatches   93;  Indels    7;  Gaps    2;

Qy     899 CAAAACAATGAAAACTTTACGAAAAATCAAAAAGTTGAAGGACTTTAA---CGTCGAGAT 955
           || |   || | |  ||| ||||| | ||||| | | | | |||||  |       ||| ||
Db      14 CATAGTTATTACACATTTGCGAAACAACAAAACGGTTAGGTACTTTCATTGGTCCGATAT 73

Qy     956 CTCTCGTAGAAAACCTCTTTTGTAAGGTTGCATACAATACTTTTTTTTCAGACTTTACTT 1015
           || | | ||||||| ||   |  ||  |   | | || ||||   |||||   |. | ||
Db      74 CTATTGGAGAAAACATCGAATTAAATTTCAAAAAAAACACTTCCCTTTCAAGATAGAATT 133

Qy    1016 ATGGTATTATACTGAATATGTTAT----TGCTGTTATAGTAGTTGAGTGACGTTTGAGGG 1071
           |  | |  |  | ||| |   ||     |  | | | | || ||  |||||  ||  ||
Db     134 TTTCTTTCGTTATTAATTTCACATAGTATTGTATCAGATTATTTTGGTGACACTTCGGGC 193

Qy    1072 AATTTCTAGTCCGTTAATCTTGTACTCAGTGTGTCTACTTT 1112
            ||   | |||   || | || |  |     ||| ||||||
Db     194 CATGCTTCGTCTAATACTTTTTTTGTGGCATTGTATACTTT 234


RESULT 13
US-08-766-439-40/c
; Sequence 40, Application US/08766439
; Patent No. 5922538
;  GENERAL INFORMATION:
;    APPLICANT:  HAZEL, JAMES WILLIAM
;    APPLICANT:  JENSEN, MARK ANTON
;    TITLE OF INVENTION:  GENETIC MARKERS AND METHODS FOR
;    TITLE OF INVENTION:  THE DETECTION OF LISTERIA
;    TITLE OF INVENTION:  MONOCYTOGENES AND LISTERIA SPP.
;    NUMBER OF SEQUENCES:  110
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  E. I. DU PONT DE NEMOURS AND COMPANY
;      STREET:  1007 MARKET STREET
;      CITY:  WILMINGTON
;      STATE:  DELAWARE
;      COUNTRY:  U.S.A.
;      ZIP:  19898
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.50 INCH DISKETTE
;      COMPUTER:  IBM PC COMPATIBLE
;      OPERATING SYSTEM:  MICROSOFT WINDOWS 3.1
;      SOFTWARE:  MICROSOFT WORD 2.0C
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/766,439
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/745,228
;      FILING DATE:  NOVEMBER 8, 1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  FLOYD, LINDA AXAMETHY
;      REGISTRATION NUMBER:  33,692
;      REFERENCE/DOCKET NUMBER:  MD-1065-A
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  302-892-8112
;      TELEFAX:  302-773-0164
;  INFORMATION FOR SEQ ID NO:  40:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1328 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    ANTI-SENSE:  YES
;    ORIGINAL SOURCE:
;      STRAIN:  L IVANOVII 3340 D.F.
US-08-766-439-40

  Query Match             3.1%;  Score 38.2;  DB 2;  Length 1328;
  Best Local Similarity   54.8%;  Pred. No. 1.3;
  Matches  121;  Conservative    0;  Mismatches   93;  Indels    7;  Gaps    2;

Qy     899 CAAAACAATGAAAACTTTACGAAAAATCAAAAAGTTGAAGGACTTTAA---CGTCGAGAT 955
           || |   || | |  ||| ||||| | ||||| | | | | |||||  |       ||| ||
Db    1315 CATAGTTATTACACATTTGCGAAACAACAAAACGGTTAGGTACTTTCATTGGTCCGATAT 1256

Qy     956 CTCTCGTAGAAAACCTCTTTTGTAAGGTTGCATACAATACTTTTTTTTCAGACTTTACTT 1015
           || | | ||||||| ||   |  ||  |   | | || ||||   |||||   |. | ||
Db    1255 CTATTGGAGAAAACATCGAATTAAATTTCAAAAAAAACACTTCCCTTTCAAGATAGAATT 1196

Qy    1016 ATGGTATTATACTGAATATGTTAT----TGCTGTTATAGTAGTTGAGTGACGTTTGAGGG 1071
           |  | |  |  | ||| |   ||     |  | | | | || ||  |||||  ||  ||
Db    1195 TTTCTTTCGTTATTAATTTCACATAGTATTGTATCAGATTATTTTGGTGACACTTCGGGC 1136

Qy    1072 AATTTCTAGTCCGTTAATCTTGTACTCAGTGTGTCTACTTT 1112
            ||   | |||   || | || |  |     ||| ||||||
Db    1135 CATGCTTCGTCTAATACTTTTTTTGTGGCATTGTATACTTT 1095


RESULT 14
US-09-071-224-3/c
; Sequence 3, Application US/09071224
; Patent No. 6271343
;  GENERAL INFORMATION:
;    APPLICANT:  Lok, Si
;    APPLICANT:  Presnell, Scott R.
;    APPLICANT:  Jelmberg, Anna C.
;    APPLICANT:  Gilbert, Teresa
;    APPLICANT:  Foster, Donald C.
;    APPLICANT:  Adams, Robyn L.
;    APPLICANT:  Lehner, Joyce M.
;    TITLE OF INVENTION:  MAMMALIAN ZCYTOR5
;    NUMBER OF SEQUENCES:  37
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Zymogenetics
;      STREET:  1201 Eastlake Ave East
;      CITY:  Seattle
;      STATE:  WA
;      COUNTRY:  USA
;      ZIP:  98102
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/071,224
;      FILING DATE:
;      CLASSIFICATION:
```

```
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lunn, Paul G
;      REGISTRATION NUMBER:  32,743
;      REFERENCE/DOCKET NUMBER:  96-22
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  206-442-6627
;      TELEFAX:  206-442-6678
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1813 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  Coding Sequence
;      LOCATION:  88...1362
;      OTHER INFORMATION:
US-09-071-224-3

  Query Match             3.1%;  Score 38.2;  DB 4;  Length 1813;
  Best Local Similarity   50.3%;  Pred. No. 1.5;
  Matches   94;  Conservative    0;  Mismatches   93;  Indels    0;  Gaps    0;

Qy     541 TTCTTTTCTTCAATTTGGTCTTGTTTTTTTTTTTTCATGATGTCATTGAATTATTCAAGA 600
           || |||| ||   ||| | || ||||||||||||  |  | |  ||    |   ||
Db    1737 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTGCACCTAAAT 1678

Qy     601 AGTCACTTCGAGCATAATGATTTTTCAAAATCCACCTTTGTTCAAGCACTACCACGTCTT 660
           ||    ||     | || || | ||    |    || | | | | |     | | ||
Db    1677 AGCTCATTTATTTAAAAGGACTCTTTTGGAGGGGCCCTAGGTAATGGGGAGTAATGACTC 1618

Qy     661 TTCATCTAGCCCACAACCGTGGTGGAGGATCTAGAATTTTCATGAAAGGATTCAAAATTT 720
             | |||  ||| || || |||  | || |||||     |||   ||   |   | |
Db    1617 CCCTTCTCACCCCCAGCCCTGGCAGGGGTTCTAGGCAACTCAACCAACCCTCACACACAC 1558

Qy     721 ACAAACA 727
           ||| |||
Db    1557 ACACACA 1551


RESULT 15
5231168-1/c
;Patent No. 5231168
;    APPLICANT: DZIEGIEL, MORTEN;BORRE, MARTIN;JEPSEN, SOREN;
;VUUST, JENS;RIENECK, KLAUS;WIND, ANNETTE;JAKOBSEN, PALLE H.
;    TITLE OF INVENTION: MALARIA ANTIGEN
;    NUMBER OF SEQUENCES: 19
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/409,658
;      FILING DATE: 18-SEP-1989
;SEQ ID NO:1:
;      LENGTH: 3095
5231168-1

  Query Match             3.1%;  Score 38;  DB 6;  Length 3095;
  Best Local Similarity   43.5%;  Pred. No. 1.9;
  Matches  173;  Conservative    0;  Mismatches  225;  Indels    0;  Gaps    0;

Qy     184 ATTGTGATTATGTCCTTCGTTGATTATTCTGTTTCATACAAGTCGTGTAATTTGCTGTTT 243
           ||  | || |    ||  ||  || ||   ||   |     || |   || |  | |||
Db    2756 ATGATAATAAAAAAATTTTTTTTTCTTTCTTTCATTTATTATCATTATATATAATATTT 2697

Qy     244 GTGACAGTACGATAGATCGACTCAACCTTCTGAGGTATTAGTTGAAGTTCATGTAAATTA 303
           | | | ||  | | |       ||  |||| |  ||||| |  ||   ||   || ||
Db    2696 CTTAAAATAATACATAAAATAATAATTTTCTTATTTATTATTATAAAGCCAGAAAATTTT 2637

Qy     304 GCTTTGTTTATCATAGTAGCATTTGATTATTGATGCTCTGTAGCTAATGATAAGCCATTG 363
             | |       ||| |  ||  | || |   || || |  |  |    |||  . |||

Db    2636 TTTATTACATGTATACTTTCAGATTATAAAATATACTGTAAATGTTTCTATATTTTATTT 2577

Qy     364 GAGGGAAGCAAGCTTTCTAAATGAATCTACGAATGGATGATAAAGTTCATGAATATTTTT 423
                 |          | | |||  | |||  |||  ||   ||     |  | | | ||||
Db    2576 ATTATATTTTTAGAATATTAATACAACTAAGAAAAGAAATTATTTATTTTTATTTTTTC 2517

Qy     424 GTTACTTCTGCAGTCAGATCATGAGTTATTGAGTCTATTGTTTTTTTAAGCCTGTTTCAG 483
             || |  |  |    ||   |   |   |   | |    |  |   |. || ||  |
Db    2516 TATAATAATATACAATGAATTTATTTCTATATATACAGAAATCCTAGTATTCTTTTATAT 2457

Qy     484 ATGATCCATCATCAGTAACAACATACACGGTGTAGTCCCAAATCCATCATATGCACCTTC 543
           |  ||  ||  | | ||| ||   | |     ||      ||     |  | |       |
Db    2456 AATATATATATTTATTAATAAATAATATTAGTTACATATAATAAGAAAAAAAATTTTATT 2397

Qy     544 TTTTCTTCAATTTGGTCTTGTTTTTTTTTTTTCATGAT 581
           |  | ||   |||  | || ||||||||||||  || |
Db    2396 TAATATTTTTTTTTTTTTTTTTTTTTTTTTTTTTTGTT 2359


Search completed: May 24, 2003, 08:06:54
Job time : 168 secs
```

GenCore version 5.1.5

OM nucleic - nucleic search, using sw model

Run on: May 24, 2003, 06:31:59 ; Search time 320 Seconds
(without alignments)
8564.631 Million cell updates/sec

Title: US-09-701-926B-1
Perfect score: 1217
Sequence: 1 tttgaaatttatgtatttat.....gtcaacacaacacaacaaca 1217

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1217	100.0	1217	21	AAZ29769	Tomato alpha-amyla
2	1217	100.0	6263	21	AAZ29771	Ds insertion mutan
3	471.4	38.7	1114	21	AAZ29770	Potato alpha-amyla
4	471.4	38.7	2094	17	AAT30126	Alpha-amylase 1 pr
5	53	4.4	8952	22	AAS46446	Tumour suppressor
6	51	4.2	6531	24	ABL32641	Human immune syste
7	49.2	4.0	5647	24	ABL70355	Chemically treated
8	49.2	4.0	5647	24	ABL33566	Human immune syste
9	49.2	4.0	5647	24	AAS61320	Human gene regulat
10	48.4	4.0	15674	24	ABL70514	Chemically treated
11	48.4	4.0	15674	24	ABL32363	Human immune syste
12	48.4	4.0	15674	24	ABL34477	Human metastasis a
13	46.2	3.8	883	22	AAL15210	Human breast cance
14	46.2	3.8	19236	24	ABN80226	Human chemically m
15	46	3.8	11805	24	ABL33748	Human immune syste
c 16	46	3.8	37515	24	ABQ66997	Human angiogenesis
17	45.8	3.8	5209	24	ABL32186	Human immune syste
18	45.8	3.8	6222	24	ABL32693	Human immune syste
19	45.6	3.7	12986	22	AAS46553	Tumour suppressor
20	45.4	3.7	8210	24	ABL70331	Chemically treated
21	45.4	3.7	8210	24	AAS61282	Human gene regulat
22	45.4	3.7	8210	24	ABK31380	Signal transductio
23	45	3.7	596	23	ABV55654	Human prostate exp
c 24	44.8	3.7	875	22	AAI95044	Human neuroblastom
25	44.8	3.7	5610	22	AAS45461	Chemically pretrea
26	44.8	3.7	5610	24	ABL92280	Chemically treated
27	44.8	3.7	5610	24	ABL49353	Human polynucleoti
28	44.8	3.7	5610	24	AAD22331	Chemically treated
29	44.8	3.7	5610	24	ABK28317	DNA transcription
30	44.8	3.7	73334	24	ABL92318	Chemically treated
31	44.8	3.7	73334	24	ABL34124	Human immune syste
32	44.6	3.7	13449	24	ABL33384	Human immune syste
33	44.6	3.7	18218	24	ABL33949	Human immune syste
34	44.4	3.6	5689	22	AAS45384	Chemically pretrea
35	44.4	3.6	5689	22	AAS46426	Tumour suppressor
36	44.4	3.6	5689	24	ABK28226	DNA transcription
37	44.4	3.6	8842	22	AAS45367	Chemically pretrea
38	44.4	3.6	8842	24	ABK28204	DNA transcription
39	44.4	3.6	8943	24	ABK39967	Human chemically p
40	44.2	3.6	6048	24	ABQ67002	Human angiogenesis
41	44.2	3.6	8897	24	ABL70228	Chemically treated
42	43.8	3.6	17703	24	ABK39953	Human chemically p
43	43.6	3.6	392	22	AAI80366	Human polynucleoti
44	43.6	3.6	6073	24	ABL33543	Human immune syste
45	43.6	3.6	6095	22	AAS46310	Tumour suppressor

ALIGNMENTS

RESULT 1
AAZ29769
ID AAZ29769 standard; DNA; 1217 BP.
XX
AC AAZ29769;
XX
DT 27-MAR-2000 (first entry)
XX
DE Tomato alpha-amylase gene promoter
XX
KW Tomato alpha-amylase promoter sequence; alpha-amylase;
KW phenotype modulating genetic sequence; PMGS; transposon tagging;
KW Ds element; dissociation element; UQ406 sequence; starch metabolism;
KW plant pathogen resistance; senescence timing; cell growth; ds.
XX
OS Lycopersicon esculentum.
XX
FH Key Location/Qualifiers
FT transposon 605..612
FT /*tag= a
FT /note= "UQ406 insertion with single Ds element"
XX
PN WO9963068-A1.
XX
PD 09-DEC-1999
XX
PF 04-JUN-1999; 99WO-AU00434.
XX
PR 04-JUN-1998; 98AU-0003901.
PR 04-JUN-1998; 98AU-0003903.
PR 25-SEP-1998; 98AU-0006169.
PR 25-SEP-1998; 98AU-0006174.

```
XX
PA   (UYQU ) UNIV QUEENSLAND.
XX
PI   Carroll BJ;
XX
DR   WPI; 2000-116368/10.
XX
PT   New polynucleotides that increase gene expression in plants used to
PT   produce transgenic plants with resistance to plant pathogens -
XX
PS   Claim 20; Fig 4; 93pp; English.
XX
CC   The present sequence is tomato alpha-amylase promoter sequence. This
CC   shows strong similarity to an alpha-amylase promoter of potato. The
CC   manipulation of amylase levels is done by introducing isolated phenotype
CC   modulating genetic sequence which increases or stabilizes expression of a
CC   second nucleotide sequence inserted proximally. GenomeWalker (14) is used
CC   to clone tomato DNA sequences flanking the Ds element in UQ406. The Ds
CC   sequence inserts into the promoter region. This is used in transposon
CC   tagging of alpha-amylase gene to identify mutants exhibiting altered
CC   physiological properties. Transgenic plants having altered phenotypic
CC   traits, such as resistance to plant pathogens, senescence timing, starch
CC   metabolism, cell growth, expansion and/or division, and the shape of
CC   cells, tissues or organs can be produced.
XX
SQ   Sequence 1217 BP; 353 A; 223 C; 200 G; 441 T; 0 other;

  Query Match             100.0%;  Score 1217;  DB 21;  Length 1217;
  Best Local Similarity   100.0%;  Pred. No. 1.1e-271;
  Matches 1217;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 TTTGAAATTTATGTATTTATCTATAGCATTAGAAACTATAAGAGTTGTTAGCTTCACTTG 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 TTTGAAATTTATGTATTTATCTATAGCATTAGAAACTATAAGAGTTGTTAGCTTCACTTG 60

Qy      61 GCTTACTGTTGTGCTCAAAGCAACTTCATCATCATACAGTATGGTTTTGATATGCTCTTC 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 GCTTACTGTTGTGCTCAAAGCAACTTCATCATCATACAGTATGGTTTTGATATGCTCTTC 120

Qy     121 CATTATCACTGAGCCTTATGATTATGTTTTACGAGCTTATAATATCACTGATGGTGATTC 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 CATTATCACTGAGCCTTATGATTATGTTTTACGAGCTTATAATATCACTGATGGTGATTC 180

Qy     181 AGTATTGTGATTATGTCCTTCGTTGATTATTCTGTTTCATACAAGTCGTGTAATTTGCTG 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 AGTATTGTGATTATGTCCTTCGTTGATTATTCTGTTTCATACAAGTCGTGTAATTTGCTG 240

Qy     241 TTTGTGACAGTACGATAGATCGACTCAACCTTCTGAGGTATTAGTTGAAGTTCATGTAAA 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 TTTGTGACAGTACGATAGATCGACTCAACCTTCTGAGGTATTAGTTGAAGTTCATGTAAA 300

Qy     301 TTAGCTTTGTTTATCATAGTAGCATTTGATTATTGATGCTCTGTAGCTAATGATAAGCCA 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 TTAGCTTTGTTTATCATAGTAGCATTTGATTATTGATGCTCTGTAGCTAATGATAAGCCA 360

Qy     361 TTGGAGGGAAGCAAGCTTTCTAAATGAATCTACGAATGGATGATAAAGTTCATGAATATT 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 TTGGAGGGAAGCAAGCTTTCTAAATGAATCTACGAATGGATGATAAAGTTCATGAATATT 420

Qy     421 TTTGTTACTTCTGCAGTCAGATCATGAGTTATTGAGTCTATTGTTTTTTTAAGCCTGTTT 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 TTTGTTACTTCTGCAGTCAGATCATGAGTTATTGAGTCTATTGTTTTTTTAAGCCTGTTT 480

Qy     481 CAGATGATCCATCATCAGTAACAACATACACGGTGTAGTCCCAAATCCATCATATGCACC 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 CAGATGATCCATCATCAGTAACAACATACACGGTGTAGTCCCAAATCCATCATATGCACC 540

Qy     541 TTCTTTTCTTCAATTTGGTCTTGTTTTTTTTTTTTCATGATGTCATTGAATTATTCAAGA 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 TTCTTTTCTTCAATTTGGTCTTGTTTTTTTTTTTTCATGATGTCATTGAATTATTCAAGA 600

Qy     601 AGTCACTTCGAGCATAATGATTTTTCAAAATCCACCTTTGTTCAAGCACTACCACGTCTT 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     601 AGTCACTTCGAGCATAATGATTTTTCAAAATCCACCTTTGTTCAAGCACTACCACGTCTT 660

Qy     661 TTCATCTAGCCCACAACCGTGGTGGAGGATCTAGAATTTTCATGAAAGGATTCAAAATTT 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     661 TTCATCTAGCCCACAACCGTGGTGGAGGATCTAGAATTTTCATGAAAGGATTCAAAATTT 720

Qy     721 ACAAACATATATATACACTATACACTATGAATCCACTAATACTAGATGGTGCACCTGTGC 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     721 ACAAACATATATATACACTATACACTATGAATCCACTAATACTAGATGGTGCACCTGTGC 780

Qy     781 CCCCACTCATGTGAAAGCCTATTCTCAATTTTTTATTTTCCACAACTTAAATACAGACCG 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     781 CCCCACTCATGTGAAAGCCTATTCTCAATTTTTTATTTTCCACAACTTAAATACAGACCG 840

Qy     841 CACAACTCCCGTGTCTTGTGTGCTCGTCGCTCAGCATGCAAGTCGAGAAAAGAAAGACCA 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     841 CACAACTCCCGTGTCTTGTGTGCTCGTCGCTCAGCATGCAAGTCGAGAAAAGAAAGACCA 900

Qy     901 AAACAATGAAAACTTTACGAAAAATCAAAAAGTTGAAGGACTTTAACGTCGAGATCTCTC 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     901 AAACAATGAAAACTTTACGAAAAATCAAAAAGTTGAAGGACTTTAACGTCGAGATCTCTC 960

Qy     961 GTAGAAAACCTCTTTTGTAAGGTTGCATACAATACTTTTTTTTCAGACTTTACTTATGGT 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     961 GTAGAAAACCTCTTTTGTAAGGTTGCATACAATACTTTTTTTTCAGACTTTACTTATGGT 1020

Qy    1021 ATTATACTGAATATGTTATTGCTGTTATAGTAGTTGAGTGACGTTTGAGGGAATTTCTAG 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1021 ATTATACTGAATATGTTATTGCTGTTATAGTAGTTGAGTGACGTTTGAGGGAATTTCTAG 1080

Qy    1081 TCCGTTAATCTTGTACTCAGTGTGTCTACTTTTCAAAAAAGTCAGTTTTTCAGTCTCTAA 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1081 TCCGTTAATCTTGTACTCAGTGTGTCTACTTTTCAAAAAAGTCAGTTTTTCAGTCTCTAA 1140

Qy    1141 AACACATTTAAATAAGAGTTTCTTTGCCCATCTTTTGTTCCTCATCCTAGGCTTGGAGTC 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1141 AACACATTTAAATAAGAGTTTCTTTGCCCATCTTTTGTTCCTCATCCTAGGCTTGGAGTC 1200

Qy    1201 AACACAACACAACAACA 1217
          |||||||||||||||||
Db    1201 AACACAACACAACAACA 1217


RESULT 2
AAZ29771
ID   AAZ29771 standard; DNA; 6263 BP.
XX
AC   AAZ29771;
XX
DT   27-MAR-2000  (first entry)
XX
DE   Ds insertion mutant of Dem coding sequence in tomato line UQ406.
XX
KW   Tomato alpha-amylase promoter sequence; alpha-amylase; Dem cDNA;
KW   Defective embryo and meristem gene; Ds element; dissociation element;
KW   phenotype modulating genetic sequence; PMGS; transposon tagging;
KW   UQ406 sequence; starch metabolism; plant pathogen resistance;
KW   senescence timing; cell growth; ds.
XX
OS   Lycopersicon esculentum.
XX
FH   Key             Location/Qualifiers
FT   transposon      644..656
FT                   /*tag= a
FT                   /note= "Ds insertion in UQ406"
FT   promoter        3388..4096
FT                   /*tag= b
FT                   /note= "Putative Dem promoter"
FT   misc_feature    4097..6263
FT                   /*tag= c
FT                   /note= "Dem cDNA coding sequence"
XX
PN   WO9963068-A1
```

```
XX
PD   09-DEC-1999.
XX
PF   04-JUN-1999;   99WO-AU00434.
XX
PR   04-JUN-1998;   98AU-0003901.
PR   04-JUN-1998;   98AU-0003903.
PR   25-SEP-1998;   98AU-0006169.
PR   25-SEP-1998;   98AU-0006174.
XX
PA   (UYQU ) UNIV QUEENSLAND.
XX
PI   Carroll BJ;
XX
DR   WPI; 2000-116368/10.
XX
PT   New polynucleotides that increase gene expression in plants used to
PT   produce transgenic plants with resistance to plant pathogens -
XX
PS   Claim 20; Fig 5; 93pp; English.
XX
CC   The present sequence is the nucleotide sequence of genomic DNA upstream
CC   of Dem gene followed by Dem cDNA coding sequence in tomato line UQ406.
CC   The manipulation of amylase levels is done by introducing isolated
CC   phenotype modulating genetic sequence which increases or stabilises
CC   expression of a second nucleotide sequence inserted proximally. The Ds
CC   sequence inserts into the promoter region. This is used in transposon
CC   tagging of alpha-amylase gene to identify mutants exhibiting altered
CC   physiological properties. Transgenic plants having altered phenotypic
CC   traits, such as resistance to plant pathogens, senescence timing, starch
CC   metabolism, cell growth, expansion and/or division, and the shape of
CC   cells, tissues or organs can be produced.
XX
SQ   Sequence 6263 BP; 1846 A; 1099 C; 1296 G; 2004 T; 18 other;

  Query Match             100.0%;  Score 1217;  DB 21;  Length 6263;
  Best Local Similarity   100.0%;  Pred. No. 1.7e-271;
  Matches 1217;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 TTTGAAATTTATGTATTTATCTATAGCATTAGAAACTATAAGAGTTGTTAGCTTCACTTG 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      40 TTTGAAATTTATGTATTTATCTATAGCATTAGAAACTATAAGAGTTGTTAGCTTCACTTG 99

Qy      61 GCTTACTGTTGTGCTCAAAGCAACTTCATCATCATACAGTATGGTTTTGATATGCTCTTC 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     100 GCTTACTGTTGTGCTCAAAGCAACTTCATCATCATACAGTATGGTTTTGATATGCTCTTC 159

Qy     121 CATTATCACTGAGCCTTATGATTATGTTTTACGAGCTTATAATATCACTGATGGTGATTC 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     160 CATTATCACTGAGCCTTATGATTATGTTTTACGAGCTTATAATATCACTGATGGTGATTC 219

Qy     181 AGTATTGTGATTATGTCCTTCGTTGATTATTCTGTTTCATACAAGTCGTGTAATTTGCTG 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     220 AGTATTGTGATTATGTCCTTCGTTGATTATTCTGTTTCATACAAGTCGTGTAATTTGCTG 279

Qy     241 TTTGTGACAGTACGATAGATCGACTCAACCTTCTGAGGTATTAGTTGAAGTTCATGTAAA 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     280 TTTGTGACAGTACGATAGATCGACTCAACCTTCTGAGGTATTAGTTGAAGTTCATGTAAA 339

Qy     301 TTAGCTTTGTTTATCATAGTAGCATTTGATTATTGATGCTCTGTAGCTAATGATAAGCCA 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     340 TTAGCTTTGTTTATCATAGTAGCATTTGATTATTGATGCTCTGTAGCTAATGATAAGCCA 399

Qy     361 TTGGAGGGAAGCAAGCTTTCTAAATGAATCTACGAATGGATGATAAAGTTCATGAATATT 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     400 TTGGAGGGAAGCAAGCTTTCTAAATGAATCTACGAATGGATGATAAAGTTCATGAATATT 459

Qy     421 TTTGTTACTTCTGCAGTCAGATCATGAGTTATTGAGTCTATTGTTTTTTTAAGCCTGTTT 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     460 TTTGTTACTTCTGCAGTCAGATCATGAGTTATTGAGTCTATTGTTTTTTTAAGCCTGTTT 519

Qy     481 CAGATGATCCATCATCAGTAACAACATACACGGTGTAGTCCCAAATCCATCATATGCACC 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     520 CAGATGATCCATCATCAGTAACAACATACACGGTGTAGTCCCAAATCCATCATATGCACC 579

Qy     541 TTCTTTTCTTCAATTTGGTCTTGTTTTTTTTTTTTTCATGATGTCATTGAATTATTCAAGA 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     580 TTCTTTTCTTCAATTTGGTCTTGTTTTTTTTTTTTTCATGATGTCATTGAATTATTCAAGA 639

Qy     601 AGTCACTTCGAGCATAATGATTTTTCAAAATCCACCTTTGTTCAAGCACTACCACGTCTT 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     640 AGTCACTTCGAGCATAATGATTTTTCAAAATCCACCTTTGTTCAAGCACTACCACGTCTT 699

Qy     661 TTCATCTAGCCCACAACCGTGGTGGAGGATCTAGAATTTTCATGAAAGGATTCAAAATTT 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     700 TTCATCTAGCCCACAACCGTGGTGGAGGATCTAGAATTTTCATGAAAGGATTCAAAATTT 759

Qy     721 ACAAACATATATATACACTATACACTATGAATCCACTAATACTAGATGGTGCACCTGTGC 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     760 ACAAACATATATATACACTATACACTATGAATCCACTAATACTAGATGGTGCACCTGTGC 819

Qy     781 CCCCACTCATGTGAAAGCCTATTCTCAATTTTTTATTTTCCACAACTTAAATACAGACCG 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     820 CCCCACTCATGTGAAAGCCTATTCTCAATTTTTTATTTTCCACAACTTAAATACAGACCG 879

Qy     841 CACAACTCCCGTGTCTTGTGTGCTCGTCGCTCAGCATGCAAGTCGAGAAAAGAAAGACCA 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     880 CACAACTCCCGTGTCTTGTGTGCTCGTCGCTCAGCATGCAAGTCGAGAAAAGAAAGACCA 939

Qy     901 AAACAATGAAAACTTTACGAAAAATCAAAAAGTTGAAGGACTTTAACGTCGAGATCTCTC 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     940 AAACAATGAAAACTTTACGAAAAATCAAAAAGTTGAAGGACTTTAACGTCGAGATCTCTC 999

Qy     961 GTAGAAAACCTCTTTTGTAAGGTTGCATACAATACTTTTTTTTCAGACTTTACTTATGGT 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1000 GTAGAAAACCTCTTTTGTAAGGTTGCATACAATACTTTTTTTTCAGACTTTACTTATGGT 1059

Qy    1021 ATTATACTGAATATGTTATTGCTGTTATAGTAGTTGAGTGACGTTTGAGGGAATTTCTAG 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1060 ATTATACTGAATATGTTATTGCTGTTATAGTAGTTGAGTGACGTTTGAGGGAATTTCTAG 1119

Qy    1081 TCCGTTAATCTTGTACTCAGTGTGTCTACTTTTCAAAAAAGTCAGTTTTTCAGTCTCTAA 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1120 TCCGTTAATCTTGTACTCAGTGTGTCTACTTTTCAAAAAAGTCAGTTTTTCAGTCTCTAA 1179

Qy    1141 AACACATTTAAATAAGAGTTTCTTTGCCCATCTTTTGTTCCTCATCCTAGGCTTGGAGTC 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1180 AACACATTTAAATAAGAGTTTCTTTGCCCATCTTTTGTTCCTCATCCTAGGCTTGGAGTC 1239

Qy    1201 AACACAACACAACAACA 1217
          |||||||||||||||||
Db    1240 AACACAACACAACAACA 1256


RESULT 3
AAZ29770
ID   AAZ29770 standard; DNA; 1114 BP.
XX
AC   AAZ29770;
XX
DT   27-MAR-2000  (first entry)
XX
DE   Potato alpha-amylase gene promoter.
XX
KW   Potato alpha-amylase promoter sequence; alpha-amylase;
KW   phenotype modulating genetic sequence; PMGS; transposon tagging;
KW   Ds element; dissociation element; UQ406 sequence; starch metabolism;
KW   plant pathogen resistance; senescence timing; cell growth; ds.
XX
OS   Solanum tuberosum.
XX
FH   Key             Location/Qualifiers
FT   transposon      647..654
FT                   /*tag= a
FT                   /note= "UQ406 insertion with single Ds element"
XX
```

PN WO9963068-A1.
XX
PD 09-DEC-1999.
XX
PF 04-JUN-1999; 99WO-AU00434.
XX
PR 04-JUN-1998; 98AU-0003901.
PR 04-JUN-1998; 98AU-0003903.
PR 25-SEP-1998; 98AU-0006169.
PR 25-SEP-1998; 98AU-0006174.
XX
PA (UYQU) UNIV QUEENSLAND.
XX
PI Carroll BJ;
XX
DR WPI; 2000-116368/10.
XX
PT New polynucleotides that increase gene expression in plants used to
PT produce transgenic plants with resistance to plant pathogens -
XX
PS Claim 20; Fig 4; 93pp; English.
XX
CC The present sequence is potato alpha-amylase promoter sequence. This
CC shows strong similarity to an alpha-amylase promoter of tomato. The
CC manipulation of amylase levels is done by introducing isolated phenotype
CC modulating genetic sequence which increases or stabilizes expression of a
CC second nucleotide sequence inserted proximally. GenomeWalker (14) is used
CC to clone tomato DNA sequences flanking the Ds element in UQ406. The Ds
CC sequence inserts into the promoter region. This is used in transposon
CC tagging of alpha-amylase gene to identify mutants exhibiting altered
CC physiological properties. Transgenic plants having altered phenotypic
CC traits, such as resistance to plant pathogens, senescence timing, starch
CC metabolism, cell growth, expansion and/or division, and the shape of
CC cells, tissues or organs can be produced.
XX
SQ Sequence 1114 BP; 326 A; 207 C; 178 G; 403 T; 0 other;

Query Match 38.7%; Score 471.4; DB 21; Length 1114;
Best Local Similarity 77.5%; Pred. No. 2.2e-99;
Matches 680; Conservative 0; Mismatches 171; Indels 26; Gaps 8;

```
Qy      1 TTTGAAATTTATGTATTTATCTATAGCATTAGAAACTATAAGAGTTGTTAGCTTCACTTG 60
          |||||||||||||||| ||||| ||||||||||||||||||||||||||||||||||||
Db      1 TTTGAAATTTATGTATATATCTGTAGCATTAGAAACTATAAGAGTTGTTAGCTTCACTTG 60

Qy     61 GCTTACTGTTGTGCTCAAAGCAACTTCATCATCATACAGTATGGTTTTGATATGCTCTTC 120
          |||| ||||||||||||||||||   ||||||||||||||||||| ||||||||||||
Db     61 TCTTATTGTTGTGCTCAAAGCAACT---TCATCATACAGTATGGTTTTTATATGCTCTTC 117

Qy    121 CATTATCACTGAGCCTTATGATTATGTTTTACGAGCTTATAATATCACTGATGGTGATTC 180
          ||||||||| || |||||||||||| | ||||||||||||||| ||||||||||||||
Db    118 CATTATCACCGAACCTTATGATTATG-TGTACGAGCTTATAATATTACTGATGGTGATTC 176

Qy    181 AGTATTGTGATTATGTCCTTCGTTGATTATTCTGTTTCATACAAGTCGTGTAATTTGCTG 240
          ||||||  |||||||| || || |||||||||||| ||| || ||||||||||||| |||||
Db    177 AGTATTATGATTATGTCCTCCATTAATTATTCTGTTTCATACAAGTCGTGTAATTTGCTG 236

Qy    241 TTTGTGACAGTACGATAGATCGACTCAACCTTCTGAGGTATTAGTTGAAGTTCATGTAAA 300
          |||||||  |||||||| || || |||||||||||| ||| || |||||||||||| |||||
Db    237 TTTGTGATTGTACGATAAATTGATTCAACCTTCTGCGGTGTTGGTTGAAGTTCAAGTAAA 296

Qy    301 TTAGCTTTGTTTATCATAGTAGCATTTGATTATTGATGCTCTGTAGCTAATGATAAGCCA 360
          ||||||||  ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    297 TTAGCTTTATTTATCATAGTAGCATTTGATTATTGATGCTCTGTAGCTAATGATAAGCCA 356

Qy    361 TTGGAGGGAAGC-----------AAGCTTTCTAAATGAATCTACGAATGGATGATAAAGT 409
          ||| ||||||||           |   ||   ||||||||||||||||||||||||||||
Db    357 TTGAAGGGAAGCAGAAATGGTAAAGCTTTCTAAAATGAATCTACGAATGGATGATAAAGT 416

Qy    410 TCATGAATATTTTTGTTACTTCTGCAGTCAGATCATGAGTTATTGAGTCTATTGTTTTTT 469
          | |||||||||  ||| ||||||||||  ||||||  |||||||  ||||||||  || |||||
Db    417 TAATGAATATTGTTGATACTTCTGCAATCAGATTATGAGTTACTGAGTCTACTG-TTTTT 475

Qy    470 TAAGCCTGTTTCAGATGATCCATCATCAGTAACAACATACACGGTGTAGTCCCAAATCCA 529
          |||||||||||||||||||| |||||||  |||||||||  | |||||||    | |
Db    476 TAAGCCTGTTTCAGATGATCGATCATCAACAACAACATATTCAGTGTAGT----AGACAT 531

Qy    530 TCATATGCACCTTCTTTTCTTCAATTTGGTCTTGTTTTTTTTTTTTTCATGATGTCATTGA 589
                ||| ||||  | ||| ||| |     | |||| |          || |    ||
Db    532 GATCGATCACTTTCTAATTTTCGATTATGCACCCTCTTTTCTCCAATTTGGTCGTCTTCT 591

Qy    590 ATTATTCAAGAAGTCACTTCGAGCATAATGATTTTTCAAAATCCACCTTTGTTCAAGCAC 649
           || |||| || ||||||    | || ||  | |         ||||| ||    |||
Db    592 TTTTTTCATGATGTCACT----GAATTATTCTCTGGTCGTCCCCACCATTCAGGAAGTCA 647

Qy    650 TACCACGTCTTTTCATCTAGCCCACAACCGTGGTGGAGGATCTAGAATTTTCATGAAAGG 709
             |  |  |  |     | |   ||    |     |      |||||||||||| ||| |
Db    648 CTTCGAGCATAATGTGAAAACATCCACATTTTTCAAATCCAGCAGAATTTTCATCAAACG 707

Qy    710 ATTCAAAATTTACAAACATATATATACACTATACACTATGAATCCACTAATACTAGATGG 769
                ||  |   |  | || ||||||||  |  | | ||||||||||||| ||||||||
Db    708 GGGTTCAACAT-TTACTACATGTATACACTCTGAAGTCTGAATCCACTAATTCTAGATGG 766

Qy    770 TGCACCTGTGCCCCCCACTCATGTGAAAGCCTATTCTCAATTTTTTTATTTTCC-ACAACTT 828
          |||| |||||||||||| | |||||||||  |||||||||||||||||||||| |||||||
Db    767 TGCATCTGTGCCCCCCACACTTGTGAAAGCTTATTCTCAATTTTTTTATTTTCCAACAACTT 826

Qy    829 AAATACAGACCGCACAACTCCCGTGTCTTGTGTGCTC 865
          ||| |||||| ||||||||||||||||||||  | ||
Db    827 GAATTCAGACCACACAACTCCCGTGTCTTGTACGGTC 863
```

RESULT 4
AAT30126
ID AAT30126 standard; DNA; 2094 BP.
XX
AC AAT30126;
XX
DT 08-JAN-1997 (first entry)
XX
DE Alpha-amylase 1 promoter.
XX
KW Alpha-amylase 1; amy 1; promoter; plant; enzyme; starch breakdown; tuber;
KW reducing sugar; potato; sprout; stem tissue; dicotyledonous plant;
KW agrobacterium; crop protein; nutrition; mammal; interferon; insulin;
KW blood factor; plasminogen activator; ss.
XX
OS Solanum tuberosum.
XX

FH	Key	Location/Qualifiers
FT	CAAT_signal	1783..1786
FT		/*tag= a
FT	CAAT_signal	1863..1866
FT		/*tag= b
FT	TATA_signal	1890..1895
FT		/*tag= c
FT	TATA_signal	1906..1911
FT		/*tag= d
FT	TATA_signal	1927..1932
FT		/*tag= e

XX
PN WO9612813-A1.
XX
PD 02-MAY-1996.
XX
PF 06-JUN-1995; 95WO-EP02195.
XX
PR 21-OCT-1994; 94GB-0021292.
XX
PA (DANI-) DANISCO AS.
XX
PI Kreiberg JD, Lund M, Pedersen HF, Pedersen RF;
XX
DR WPI; 1996-230612/23.
XX
PT Plant alpha-amylase gene promoter - capable of expressing a gene of

```
PT   interest in sprout or stem tissue of a dicotyledonous plant
XX
PS   Claim 1; Page 33-36; 48pp; English.
XX
CC   This sequence represents the Solanum tuberosum alpha-amylase 1 promoter.
CC   Alpha-amylase is one of the key plant enzymes, and participates in the
CC   pathway responsible for the breakdown of starch to reducing sugars in
CC   potato tubers.  Alpha-amylase is encoded by a gene family consisting of
CC   at least five individual genes divided into two sub families (the type 1
CC   and type 3 alpha amylases).  Type 1 alpha-amylases are expressed in
CC   sprout and stem tissues, but not in tubers, roots or leaves.  The type 3
CC   alpha amylases are expressed in tubers, sprouts and stem tissues.  The
CC   promoter is used to direct expression of a gene of interest in stem
CC   cells, tissues or organs of a dicotyledonous plant (such as a potato).
CC   This sequence can be used in constructs, such as vectors (such as those
CC   derived from agrobacterium), for transforming cells to express a gene of
CC   interest.  The gene of interest that is fused to this promoter sequence
CC   may be a crop protein gene with an optimised amino acid composition, so
CC   as to increase the nutritive value of the crop.  It is also possible to
CC   use this promoter to express non-plant genes for mammalian products, such
CC   as interferons, insulin, blood factors, and plasminogen activators.
XX
SQ   Sequence 2094 BP; 618 A; 357 C; 369 G; 750 T; 0 other;

  Query Match             38.7%;  Score 471.4;  DB 17;  Length 2094;
  Best Local Similarity   77.5%;  Pred. No. 2.5e-99;
  Matches  680;  Conservative    0;  Mismatches  171;  Indels   26;  Gaps    8;

Qy       1 TTTGAAATTTATGTATTTATCTATAGCATTAGAAACTATAAGAGTTGTTAGCTTCACTTG 60
Db     981 TTTGAAATTTATGTATATATCTGTAGCATTAGAAACTATAAGAGTTGTTAGCTTCACTTG 1040

Qy      61 GCTTACTGTTGTGCTCAAAGCAACTTCATCATCATACAGTATGGTTTTGATATGCTCTTC 120
Db    1041 TCTTATTGTTGTGCTCAAAGCAACT---TCATCATACAGTATGGTTTTTATATGCTCTTC 1097

Qy     121 CATTATCACTGAGCCTTATGATTATGTTTTACGAGCTTATAATATCACTGATGGTGATTC 180
Db    1098 CATTATCACCGAACCTTATGATTATG-TGTACGAGCTTATAATATTACTGATGGTGATTC 1156

Qy     181 AGTATTGTGATTATGTCCTTCGTTGATTATTCTGTTTCATACAAGTCGTGTAATTTGCTG 240
Db    1157 AGTATTATGATTATGTCCTCCATTAATTATTCTGTTTCATACAAGTCGTGTAATTTGCTG 1216

Qy     241 TTTGTGACAGTACGATAGATCGACTCAACCTTCTGAGGTATTAGTTGAAGTTCATGTAAA 300
Db    1217 TTTGTGATTGTACGATAAATTGATTCAACCTTCTGCGGTGTTGGTTGAAGTTCAAGTAAA 1276

Qy     301 TTAGCTTTGTTTATCATAGTAGCATTTGATTATTGATGCTCTGTAGCTAATGATAAGCCA 360
Db    1277 TTAGCTTTATTTATCATAGTAGCATTTGATTATTGATGCTCTGTAGCTAATGATAAGCCA 1336

Qy     361 TTGGAGGGAAGC-----------AAGCTTTCTAAATGAATCTACGAATGGATGATAAAGT 409
Db    1337 TTGAAGGGAAGCAGAAATGGTAAAGCTTTCTAAAATGAATCTACGAATGGATGATAAAGT 1396

Qy     410 TCATGAATATTTTTGTTACTTCTGCAGTCAGATCATGAGTTATTGAGTCTATTGTTTTTT 469
Db    1397 TAATGAATATTGTTGATACTTCTGCAATCAGATTATGAGTTACTGAGTCTACTG-TTTTT 1455

Qy     470 TAAGCCTGTTTCAGATGATCCATCATCAGTAACAACATACACGGTGTAGTCCCAAATCCA 529
Db    1456 TAAGCCTGTTTCAGATGATCGATCATCAACAACAACATATTCAGTGTAGT----AGACAT 1511

Qy     530 TCATATGCACCTTCTTTTCTTCAATTTGGTCTTGTTTTTTTTTTTTTCATGATGTCATTGA 589
Db    1512 GATCGATCACTTTCTAATTTTCGATTATGCACCCTCTTTTCTCCAATTTGGTCGTCTTCT 1571

Qy     590 ATTATTCAAGAAGTCACTTCGAGCATAATGATTTTTCAAAATCCACCTTTGTTCAAGCAC 649
Db    1572 TTTTTTCATGATGTCACT----GAATTATTCTCTGGTCGTCCCCACCATTCAGGAAGTCA 1627

Qy     650 TACCACGTCTTTTCATCTAGCCCACAACCGTGGTGGAGGATCTAGAATTTTCATGAAAGG 709
```

```
Db    1628 CTTCGAGCATAATGTGAAAACATCCACATTTTTCAAATCCAGCAGAATTTTCATCAAACG 1687

Qy     710 ATTCAAAATTTACAAACATATATATACACTATACACTATGAATCCACTAATACTAGATGG 769
Db    1688 GGGTTCAACAT-TTACTACATGTATACACTCTGAAGTCTGAATCCACTAATTCTAGATGG 1746

Qy     770 TGCACCTGTGCCCCCACTCATGTGAAAGCCTATTCTCAATTTTTTATTTTCC-ACAACTT 828
Db    1747 TGCATCTGTGCCCCCACACTTGTGAAAGCTTATTCTCAATTTTTTATTTTCCAACAACTT 1806

Qy     829 AAATACAGACCGCACAACTCCCGTGTCTTGTGTGCTC 865
Db    1807 GAATTCAGACCACACAACTCCCGTGTCTTGTACGGTC 1843


RESULT 5
AAS46446
ID   AAS46446 standard; DNA; 8952 BP.
XX
AC   AAS46446;
XX
DT   18-DEC-2001  (first entry)
XX
DE   Tumour suppressor gene derived chemically modified sequence #168.
XX
KW   Human; tumour suppressor gene; oncogene; antitumour; cytostatic;
KW   cancer; tumour; CpG dinucleotide; single-nucleotide polymorphism; SNP;
KW   cytosine methylation; ds.
XX
OS   Homo sapiens.
XX
PN   WO200168912-A2.
XX
PD   20-SEP-2001.
XX
PF   15-MAR-2001; 2001WO-EP02955.
XX
PR   15-MAR-2000; 2000DE-1013847.
PR   06-APR-2000; 2000DE-1019058.
PR   07-APR-2000; 2000DE-1019173.
PR   30-JUN-2000; 2000DE-1032529.
PR   01-SEP-2000; 2000DE-1043826.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K;
XX
DR   WPI; 2001-602752/68.
XX
PT   Fragments of chemically modified genes associated with tumour suppressor
PT   genes and oncogenes, useful in designing primers and probes for
PT   analysing diseases associated with cytosine methylation state e.g.
PT   cancer -
XX
PS   Claim 1; SEQ ID No 168; 27pp; English.
XX
CC   The invention relates to a nucleic acid comprising a sequence of 18
CC   bases, of a segment of chemically pretreated DNA (CP DNA) e.g. with
CC   bisulphite, of genes associated with tumour suppression and
CC   oncogenes having a sequence taken from 536 (actually 533 since
CC   numbers 408, 458 and 500 are missing from the sequence listing) sequences
CC   (Ss) and sequences complementary to (Ss). The nucleic acid may be a
CC   peptide nucleic acid-oligomer (PNA) of at least 9 nucleotides and may
CC   form part of a set of probes for detecting the cytosine methylation state
CC   and/or single nucleotide polymorphisms and also to be used in an
CC   array for analysing diseases associated with CpG dinucleotides e.g.
CC   cancers and tumours. The probes can also be used in a method for
CC   ascertaining genetic and/or epigenetic parameters for the diagnosis
CC   and/or therapy of existing diseases or the predisposition to specific
CC   diseases, by analysing cytosine methylations. The parameters may be
CC   compared to another set of genetic and/or epigenetic parameters, the
CC   differences serving as basis for diagnosis and/or prognosis events which
```

CC are disadvantageous to patients. The present sequence is one of the
CC 533 genomic sequences derived from tumour suppressor genes and
CC oncogenes. Sequences with even numbered Seq ID numbers are the
CC complementary sequence of the corresponding odd numbered sequence (e.g.
CC ID 2 and ID1, ID 536 and ID 535, except for those whose partner sequence
CC is missing).
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 8952 BP; 2404 A; 162 C; 1905 G; 4481 T; 0 other;

Query Match 4.4%; Score 53; DB 22; Length 8952;
Best Local Similarity 50.2%; Pred. No. 0.017;
Matches 157; Conservative 0; Mismatches 155; Indels 1; Gaps 1;

```
Qy      40 AAGAGTTGTTAGCTTCACTTGGCTTACTGTTGTGCTCAAAGCAACTTCATCATCATACAG 99
           |||| || |||  ||   |||| ||  | ||||| |   || || ||  | ||  ||
Db    6392 AAGATTTTTTATTTTTTTTTTGGGTTTGTTTTGTGTTTTTAGGAATTTGTTTATGTTATTT 6451

Qy     100 TATGGTTTTGATATGCTCTTCCATTATCACTGAGCCTTATGATTATGTTTTACGAGCTTA 159
             |  ||| || || |  |   | ||   | |   ||||  |||| |  ||     |
Db    6452 AGAGTATTTAATTTGTTGGTATTTAATTGTTTA-TATTATTTTTATATAATATTTTTTAT 6510

Qy     160 TAATATCACTGATGGTGATTCAGTATTGTGATTATGTCCTTCGTTGATTATTCTGTTTCA 219
           |  |||     | ||||| |   || || |  ||   |       ||  ||||| | |||
Db    6511 TTTTATATAGGTTGGTGGTAATGTGTTTTTTTTTTATTTTGATTTTAGTTATTTTATTTTT 6570

Qy     220 TACAAGTCGTGTAATTTGCTGTTTGTGACAGTACGATAGATCGACTCAACCTTCTGAGGT 279
           |      | |||  ||||  |  ||        | | ||| |    | ||  || | | ||
Db    6571 TTGTTTTTTTGTTGTTTGTTTGTTTGTTTTATTGGTTAGTTTAGTTAAAGGTTTTTAAGT 6630

Qy     280 ATTAGTTGAAGTTCATGTAAATTAGCTTTGTTTATCATAGTAGCATTTGATTATTGATGC 339
           || ||  |  |  |  || || | ||| |||||  |  | |  |||  || ||  |
Db    6631 TTTGGTGATAAATATTAAAATTTGGTTTTATTTATTTTTTTTGGTTTTTTTTTTTTTTTT 6690

Qy     340 TCTGTAGCTAATG 352
           |  |  |  || ||
Db    6691 TTTAGATTTATTG 6703
```

RESULT 6
ABL32641
ID ABL32641 standard; DNA; 6531 BP.
XX
AC ABL32641;
XX
DT 26-MAR-2002 (first entry)
XX
DE Human immune system associated gene SEQ ID NO: 614.
XX
KW Human; immune system disease; cytosine methylation; antiasthmatic;
KW antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
KW gene; ds.
XX
OS Homo sapiens.
XX
PN WO200200928-A2.
XX
PD 03-JAN-2002.
XX
PF 02-JUL-2001; 2001WO-EP07537.
XX
PR 30-JUN-2000; 2000DE-1032529.
PR 01-SEP-2000; 2000DE-1043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-130909/17.
XX
PT Nucleic acid comprising fragment of chemically modified gene, useful
PT for diagnosis and treatment of diseases associated with abnormal
PT cytosine methylation -
XX
PS Claim 1; SEQ ID NO 614; 32pp + Sequence Listing; German.
XX
CC The present invention provides a number of human immune system associated
CC genes which are modified by the methylation of cytosines. The sequences
CC can be used in the diagnosis and treatment of immune system disorders,
CC including eye diseases such as retinopathy, neovascular glaucoma and
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC diseases. The present sequence is a gene of the invention.
XX
SQ Sequence 6531 BP; 1742 A; 131 C; 1301 G; 3357 T; 0 other;

Query Match 4.2%; Score 51; DB 24; Length 6531;
Best Local Similarity 44.2%; Pred. No. 0.047;
Matches 210; Conservative 0; Mismatches 265; Indels 0; Gaps 0;

```
Qy     157 TTATAATATCACTGATGGTGATTCAGTATTGTGATTATGTCCTTCGTTGATTATTCTGTT 216
           ||||  | |   | | |||||||  | ||| |     |     ||  ||||| ||
Db    5002 TTATTTTGTTGTTTAGGGTGATTTTGAATTTTTGCGTTTAAGTAGTTTTTTTATTTTGGA 5061

Qy     217 TCATACAAGTCGTGTAATTTGCTGTTTGTGACAGTACGATAGATCGACTCAACCTTCTGA 276
           |  |  ||||  ||  |||    ||||| |                |  | |  |  | | |
Db    5062 TTTTTAAAGTGTTGGGATTATAAGTTTGAGTTTTATTTTTTTTTTTTAATTGTTTTAGTTA 5121

Qy     277 GGTATTAGTTGAAGTTCATGTAAATTAGCTTTGTTTATCATAGTAGCATTTGATTATTGA 336
             | ||| ||    ||  | |   ||  ||| |||    ||||      | |  ||| |
Db    5122 TTTTTTAATTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTGATTTAGTTATTGGTAAAAATTTA 5181

Qy     337 TGCTCTGTAGCTAATGATAAGCCATTGGAGGGAAGCAAGCTTTCTAAATGAATCTACGAA 396
           | | | |       || | |   || ||||    |  || ||| ||  | ||     ||
Db    5182 GGTTTTTTTTTATTTGTTTATTTATAGGAGAATAATTAGTTTTTTATTTTAAGGAGGAAA 5241

Qy     397 TGGATGATAAAGTTCATGAATATTTTTGTTACTTCTGCAGTCAGATCATGAGTTATTGAG 456
           | |       | || |    | ||||| ||     |  |||   ||  | | ||  |
Db    5242 TAGTATGATTATTTTAATTTTTTTTTTTTTTAAGTTTTAGTTTAATAGTAATTTTATTTA 5301

Qy     457 TCTATTGTTTTTTTAAGCCTGTTTCAGATGATCCATCATCAGTAACAACATACACGGTGT 516
           | | |  |||||||      |||||  |||          |  |    ||      |  | |||
Db    5302 TTTTTATTTTTTTTTTTATTTGTTTTAGAAATAAAAGTAATTTTATTTTTTTTTAAGGTAA 5361

Qy     517 AGTCCCAAATCCATCATATGCACCTTCTTTTCTTCAATTTGGTCTTGTTTTTTTTTTTTC 576
           | |       |  | |              || ||||     |||  |   | |  |||  || |
Db    5362 ATTTTTGTTTAGGTTAGTATGTTGTTTTTTTTAAATTTTTTGAGATTTGGTTTAATTATT 5421

Qy     577 ATGATGTCATTGAATTATTCAAGAAGTCACTTCGAGCATAATGATTTTTCAAAAT 631
           |  || |  ||   |  ||  ||   | | ||   |  | |  || ||   |||
Db    5422 TTTATTTTTTTTTGTATTTTTAGTTTTTAGTTTTTGAGGATTTTTTATTTTTAAT 5476
```

RESULT 7
ABL70355
ID ABL70355 standard; DNA; 5647 BP.
XX
AC ABL70355;
XX
DT 01-JUL-2002 (first entry)
XX
DE Chemically treated cell signalling DNA sequence#123.
XX
KW Cell signalling; cytosine methylation; cell signalling disease;
KW cancer; tumour; cytostatic; ds.
XX
OS Unidentified.

XX
PN WO200202807-A2.
XX
PD 10-JAN-2002.
XX
PF 29-JUN-2001; 2001WO-EP07471.
XX
PR 30-JUN-2000; 2000DE-1032529.
PR 01-SEP-2000; 2000DE-1043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-154758/20.
XX
PT Nucleic acid, useful for diagnosis and therapy of diseases associated
PT with cell signalling e.g. cancer, comprises chemically modified genomic
PT sequences of genes associated with cell signalling -
XX
PS Claim 1; SEQ ID NO 245; 24pp+sequence listing; English.
XX
CC The invention relates to a nucleic acid comprising a sequence of at least
CC 18 bases of a segment of chemically pretreated DNA of genes associated
CC with cell signalling. The activity of the modified sequences of the
CC invention may be described as cytostatic. The object of the invention is
CC to provide the chemically modified DNA of genes associated with cell
CC signalling, as well as oligonucleotides and/or PNA-oligomers for
CC detecting cytosine methylations, as well as a method which is
CC particularly suitable for the diagnosis and/or therapy of genetic and
CC epigenetic parameters of genes associated with cell signalling. The
CC chemically modified DNA provided by the invention is useful for diagnosis
CC and therapy of diseases such as solid tumours and cancer. The sequences
CC given in records ABL70111-ABL70626 represent chemically pre-treated
CC genomic DNA's of genes associated with cell signalling.
CC Note: The sequence data for this patent is not represented in the printed
CC specification, but is based on sequence information supplied by the
CC European Patent Office.
XX
SQ Sequence 5647 BP; 1448 A; 52 C; 945 G; 3202 T; 0 other;

Query Match 4.0%; Score 49.2; DB 24; Length 5647;
Best Local Similarity 44.0%; Pred. No. 0.12;
Matches 207; Conservative 0; Mismatches 263; Indels 0; Gaps 0;

Qy 105 TTTTGATATGCTCTTCCATTATCACTGAGCCTTATGATTATGTTTTACGAGCTTATAATA 164
Db 2825 TTTTGATTTAATTTTAGTAAATAATATTGTATTTTTAATTTATTTTTTTTTGTGTTTTTT 2884

Qy 165 TCACTGATGGTGATTCAGTATTGTGATTATGTCCTTCGTTGATTATTCTGTTTCATACAA 224
Db 2885 TTTTTTTTTTTATTAATTTTTTTTTTATAGATTAAGTTTTTTTTTAAAAATTTTTTTTGT 2944

Qy 225 GTCGTGTAATTTGCTGTTTGTGACAGTACGATAGATCGACTCAACCTTCTGAGGTATTAG 284
Db 2945 ATTTTTTATTCGGATGTTTATGTTTATTTGGTAGTTGTTTTTAAATTTATAATAAATATT 3004

Qy 285 TTGAAGTTCATGTAAATTAGCTTTGTTTATCATAGTAGCATTTGATTATTGATGCTCTGT 344
Db 3005 TTTGAGTAAAGTAAAATTAATTAATTTTTTTTTCGTTTTTTTGAATAATGAAATGATTTT 3064

Qy 345 AGCTAATGATAAGCCATTGGAGGGAAGCAAGCTTTCTAAATGAATCTACGAATGGATGAT 404
Db 3065 AGAATTTTATTATTGAATATAAAAGTAGTATATTTTAGATTATTTTTTTATATTTGTTTT 3124

Qy 405 AAAGTTCATGAATATTTTTGTTACTTCTGCAGTCAGATCATGAGTTATTGAGTCTATTGT 464
Db 3125 TTAGTTTATTAATATTTTTTTTAGTTTTGTTTAATTTGTTATTTAATTTATGTTTTGAGT 3184

Qy 465 TTTTTTAAGCCTGTTTCAGATGATCCATCATCAGTAACAACATACACGGTGTAGTCCCAA 524
Db 3185 TTTTAAAATTTTTAATTAATTTTTTTTTTTAATTTTAGAAGTTTTATTTGGTTTTTTAAA 3244

Qy 525 ATCCATCATATGCACCTTCTTTTCTTCAATTTGGTCTTGTTTTTTTTTTTT 574
Db 3245 AAAAATTATTTAGTTTTTTTTTTTTTGATAATGTTTTATTTTTTTTTTAT 3294

RESULT 8
ABL33566
ID ABL33566 standard; DNA; 5647 BP.
XX
AC ABL33566;
XX
DT 26-MAR-2002 (first entry)
XX
DE Human immune system associated gene SEQ ID NO: 1539.
XX
KW Human; immune system disease; cytosine methylation; antiasthmatic;
KW antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
KW gene; ds.
XX
OS Homo sapiens.
XX
PN WO200200928-A2.
XX
PD 03-JAN-2002.
XX
PF 02-JUL-2001; 2001WO-EP07537.
XX
PR 30-JUN-2000; 2000DE-1032529.
PR 01-SEP-2000; 2000DE-1043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-130909/17.
XX
PT Nucleic acid comprising fragment of chemically modified gene, useful
PT for diagnosis and treatment of diseases associated with abnormal
PT cytosine methylation -
XX
PS Claim 1; SEQ ID NO 1539; 32pp + Sequence Listing; German.
XX
CC The present invention provides a number of human immune system associated
CC genes which are modified by the methylation of cytosines. The sequences
CC can be used in the diagnosis and treatment of immune system disorders,
CC including eye diseases such as retinopathy, neovascular glaucoma and
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC diseases. The present sequence is a gene of the invention.
XX
SQ Sequence 5647 BP; 1448 A; 52 C; 945 G; 3202 T; 0 other;

Query Match 4.0%; Score 49.2; DB 24; Length 5647;
Best Local Similarity 44.0%; Pred. No. 0.12;
Matches 207; Conservative 0; Mismatches 263; Indels 0; Gaps 0;

Qy 105 TTTTGATATGCTCTTCCATTATCACTGAGCCTTATGATTATGTTTTACGAGCTTATAATA 164
Db 2825 TTTTGATTTAATTTTAGTAAATAATATTGTATTTTTAATTTATTTTTTTTTGTGTTTTTT 2884

Qy 165 TCACTGATGGTGATTCAGTATTGTGATTATGTCCTTCGTTGATTATTCTGTTTCATACAA 224
Db 2885 TTTTTTTTTTTATTAATTTTTTTTTTATAGATTAAGTTTTTTTTTAAAAATTTTTTTTGT 2944

Qy 225 GTCGTGTAATTTGCTGTTTGTGACAGTACGATAGATCGACTCAACCTTCTGAGGTATTAG 284
Db 2945 ATTTTTTATTCGGATGTTTATGTTTATTTGGTAGTTGTTTTTAAATTTATAATAAATATT 3004

```
Qy     285 TTGAAGTTCATGTAAATTAGCTTTGTTTATCATAGTAGCATTTGATTATTGATGCTCTGT 344
           ||  |||  |   ||||||  |   ||| |  | ||    ||  || ||  |     | |
Db    3005 TTTGAGTAAAGTAAAATTAATTAATTTTTTTTTCGTTTTTTTGAATAATGAAATGATTTT 3064

Qy     345 AGCTAATGATAAGCCATTGGAGGGAAGCAAGCTTTCTAAATGAATCTACGAATGGATGAT 404
           ||    | || |   | | |         |  |||   | |   | |    ||  |  |
Db    3065 AGAATTTTATTATTGAATATAAAAGTAGTATATTTTAGATTATTTTTTTATATTTGTTTT 3124

Qy     405 AAAGTTCATGAATATTTTTGTTACTTCTGCAGTCAGATCATGAGTTATTGAGTCTATTGT 464
             |||| ||  |||||||||  ||| || ||                ||  || |   ||
Db    3125 TTAGTTTATTAATATTTTTTTTAGTTTTGTTTAATTTGTTATTTAATTTATGTTTTGAGT 3184

Qy     465 TTTTTTAAGCCTGTTTCAGATGATCCATCATCAGTAACAACATACACGGTGTAGTCCCAA 524
           ||||  ||   |   | |  |  |   |      | | ||  |  |    ||  |   ||
Db    3185 TTTTAAAATTTTTAATTAATTTTTTTTTTTAATTTTAGAAGTTTTATTTGGTTTTTTAAA 3244

Qy     525 ATCCATCATATGCACCTTCTTTTCTTCAATTTGGTCTTGTTTTTTTTTTTT 574
           |   || || |     || |||| || |   || | |  |||||||||  |
Db    3245 AAAAATTATTTAGTTTTTTTTTTTTTTGATAATGTTTTATTTTTTTTTTAT 3294
```

RESULT 9
AAS61320
ID AAS61320 standard; DNA; 5647 BP.
XX
AC AAS61320;
XX
DT 29-JAN-2002 (first entry)
XX
DE Human gene regulation-associated gene oligonucleotide #275.
XX
KW Human; Gene regulation-associated gene; severe combined immunodeficiency;
KW cardiac damage; inflammatory response; Haemophilia; Werner syndrome;
KW asthma; HDR syndrome; congenital heart defect; Saethre-Chotzen syndrome;
KW renal disease; Preeclampsia; cardiac allograft vascular disease;
KW colorectal cancer; thyroid cancer; oesophageal cancer; ds; tumour;
KW immunostimulant; cardiant; antiinflammatory; coagulant; antiasthmatic;
KW nephrotropic; gynecological; anti-tumour; immunosuppressive; cytostatic.
XX
OS Homo sapiens.
XX
PN WO200177375-A2.
XX
PD 18-OCT-2001.
XX
PF 06-APR-2001; 2001WO-EP03968.
XX
PR 06-APR-2000; 2000DE-1019058.
PR 07-APR-2000; 2000DE-1019173.
PR 30-JUN-2000; 2000DE-1032529.
PR 01-SEP-2000; 2000DE-1043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-017470/02.
XX
PT New nucleic acid sequences from chemically modified genes associated
PT with gene regulation, useful for analysing cytosine methylations for
PT diagnosis and therapy of diseases e.g. severe combined immunodeficiency
PT disease -
XX
PS Disclosure; SEQ ID No 281; 26pp; English.
XX
CC The invention relates to 224 nucleic acid sequences comprising at least
CC 18 bases of a chemically pretreated gene associated with gene regulation
CC selected from 43 known genes (or complementary sequences). The
CC chemical pretreatment converts cytosine bases unmethylated at the
CC 5-position to uracil or another base with hybridisation behaviour
CC dissimilar to cytosine, to enable analysis of cytosine methylations.
CC The DNA sequences, oligomers (or sets/arrays) and method are
CC useful in the diagnosis of diseases (or predisposition to diseases)
CC associated with gene regulation and in therapy of such diseases, by
CC enabling analysis of the cytosine methylation patterns of such genes,
CC kits are provided. They are especially useful in diagnosis
CC and therapy of e.g. severe combined immunodeficiency disease, cardiac
CC disorders, haemophilia, solid tumours and cancer, Werner syndrome,
CC asthma, HDR syndrome, Saethre-Chotzen syndrome, renal disease,
CC preeclampsia, graft versus-host disease. The present sequence is a
CC sequence included in the sequence data for this specification and is
CC associated with the human gene regulation-associated genes.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 5647 BP; 1448 A; 52 C; 945 G; 3202 T; 0 other;

Query Match 4.0%; Score 49.2; DB 24; Length 5647;
Best Local Similarity 44.0%; Pred. No. 0.12;
Matches 207; Conservative 0; Mismatches 263; Indels 0; Gaps 0;

```
Qy     105 TTTTGATATGCTCTTCCATTATCACTGAGCCTTATGATTATGTTTTACGAGCTTATAATA 164
           |||||||  |  | ||      || |    |  || | | | | ||||      |  |  |
Db    2825 TTTTGATTTAATTTTAGTAAATAATATTGTATTTTTAATTTATTTTTTTTTGTGTTTTTT 2884

Qy     165 TCACTGATGGTGATTCAGTATTGTGATTATGTCCTTCGTTGATTATTCTGTTTCATACAA 224
           |   |  |  |  |     | || |   ||  |      ||  ||      |||  |
Db    2885 TTTTTTTTTTTATTAATTTTTTTTTTATAGATTAAGTTTTTTTTTAAAAATTTTTTTTGT 2944

Qy     225 GTCGTGTAATTTGCTGTTTGTGACAGTACGATAGATCGACTCAACCTTCTGAGGTATTAG 284
           |  | || |  | ||||| ||    |  | ||| |    | ||  || | |   ||
Db    2945 ATTTTTTATTCGGATGTTTATGTTTATTTGGTAGTTGTTTTTAAATTTATAATAAATATT 3004

Qy     285 TTGAAGTTCATGTAAATTAGCTTTGTTTATCATAGTAGCATTTGATTATTGATGCTCTGT 344
           ||  |||  |   ||||||  |   ||| |  | ||    ||  || ||  |     | |
Db    3005 TTTGAGTAAAGTAAAATTAATTAATTTTTTTTTCGTTTTTTTGAATAATGAAATGATTTT 3064

Qy     345 AGCTAATGATAAGCCATTGGAGGGAAGCAAGCTTTCTAAATGAATCTACGAATGGATGAT 404
           ||    | || |   | | |         |  |||   | |   | |    ||  |  |
Db    3065 AGAATTTTATTATTGAATATAAAAGTAGTATATTTTAGATTATTTTTTTATATTTGTTTT 3124

Qy     405 AAAGTTCATGAATATTTTTGTTACTTCTGCAGTCAGATCATGAGTTATTGAGTCTATTGT 464
             |||| ||  |||||||||  ||| || ||                ||  || |   ||
Db    3125 TTAGTTTATTAATATTTTTTTTAGTTTTGTTTAATTTGTTATTTAATTTATGTTTTGAGT 3184

Qy     465 TTTTTTAAGCCTGTTTCAGATGATCCATCATCAGTAACAACATACACGGTGTAGTCCCAA 524
           ||||  ||   |   | |  |  |   |      | | ||  |  |    ||  |   ||
Db    3185 TTTTAAAATTTTTAATTAATTTTTTTTTTTAATTTTAGAAGTTTTATTTGGTTTTTTAAA 3244

Qy     525 ATCCATCATATGCACCTTCTTTTCTTCAATTTGGTCTTGTTTTTTTTTTT 574
           |   || || |     || |||| || |   || | |  |||||||||  |
Db    3245 AAAAATTATTTAGTTTTTTTTTTTTTTGATAATGTTTTATTTTTTTTTTAT 3294
```

RESULT 10
ABL70514
ID ABL70514 standard; DNA; 15674 BP.
XX
AC ABL70514;
XX
DT 01-JUL-2002 (first entry)
XX
DE Chemically treated cell signalling DNA sequence complementary to#202.
XX
KW Cell signalling; cytosine methylation; cell signalling disease;
KW cancer; tumour; cytostatic; ds.
XX
OS Unidentified.
XX
PN WO200202807-A2.
XX
PD 10-JAN-2002.
XX

PF 29-JUN-2001; 2001WO-EP07471.
XX
PR 30-JUN-2000; 2000DE-1032529.
PR 01-SEP-2000; 2000DE-1043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-154758/20.
XX
PT Nucleic acid, useful for diagnosis and therapy of diseases associated
PT with cell signalling e.g. cancer, comprises chemically modified genomic
PT sequences of genes associated with cell signalling -
XX
PS Claim 1; SEQ ID NO 404; 24pp+sequence listing; English.
XX
CC The invention relates to a nucleic acid comprising a sequence of at least
CC 18 bases of a segment of chemically pretreated DNA of genes associated
CC with cell signalling. The activity of the modified sequences of the
CC invention may be described as cytostatic. The object of the invention is
CC to provide the chemically modified DNA of genes associated with cell
CC signalling, as well as oligonucleotides and/or PNA-oligomers for
CC detecting cytosine methylations, as well as a method which is
CC particularly suitable for the diagnosis and/or therapy of genetic and
CC epigenetic parameters of genes associated with cell signalling. The
CC chemically modified DNA provided by the invention is useful for diagnosis
CC and therapy of diseases such as solid tumours and cancer. The sequences
CC given in records ABL70111-ABL70626 represent chemically pre-treated
CC genomic DNA's of genes associated with cell signalling.
CC Note: The sequence data for this patent is not represented in the printed
CC specification, but is based on sequence information supplied by the
CC European Patent Office.
XX
SQ Sequence 15674 BP; 4711 A; 271 C; 3032 G; 7660 T; 0 other;

Query Match 4.0%; Score 48.4; DB 24; Length 15674;
Best Local Similarity 48.2%; Pred. No. 0.23;
Matches 136; Conservative 0; Mismatches 146; Indels 0; Gaps 0;

Qy 185 TTGTGATTATGTCCTTCGTTGATTATTCTGTTTCATACAAGTCGTGTAATTTGCTGTTTG 244
Db 13049 TTTGGTTTTTATTTTTTTTTTTTTTTATTTTATAATAGTAAAGTTGTTATAGTTTTTTTT 13108

Qy 245 TGACAGTACGATAGATCGACTCAACCTTCTGAGGTATTAGTTGAAGTTCATGTAAATTAG 304
Db 13109 TTATAGATAAATAAATTTAAAAAAATTATTTTTGTTATTTTTATAGATAAGGAAAGTGAG 13168

Qy 305 CTTTGTTTATCATAGTAGCATTTGATTATTGATGCTCTGTAGCTAATGATAAGCCATTGG 364
Db 13169 GTTTGGTTATAAGTTTAGGGAGGGACGGTGGTTGGATTGAAGTTTTTTTGATTTTATGGT 13228

Qy 365 AGGGAAGCAAGCTTTCTAAATGAATCTACGAATGGATGATAAAGTTCATGAATATTTTTG 424
Db 13229 TAGCGAAAAAATTTATAAATTTTTTTTTTTAATGTATAAAAAATTTTTTTTGTTTTTATT 13288

Qy 425 TTACTTCTGCAGTCAGATCATGAGTTATTGAGTCTATTGTTT 466
Db 13289 TTTGTTTTAACGATATATTTATTTTTATTTTTTTTTTTATAT 13330

RESULT 11
ABL32363
ID ABL32363 standard; DNA; 15674 BP.
XX
AC ABL32363;
XX
DT 26-MAR-2002 (first entry)
XX
DE Human immune system associated gene SEQ ID NO: 336.
XX
KW Human; immune system disease; cytosine methylation; antiasthmatic;
KW antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
KW gene; ds.
XX
OS Homo sapiens.
XX
PN WO200200928-A2.
XX
PD 03-JAN-2002.
XX
PF 02-JUL-2001; 2001WO-EP07537.
XX
PR 30-JUN-2000; 2000DE-1032529.
PR 01-SEP-2000; 2000DE-1043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-130909/17.
XX
PT Nucleic acid comprising fragment of chemically modified gene, useful
PT for diagnosis and treatment of diseases associated with abnormal
PT cytosine methylation -
XX
PS Claim 1; SEQ ID NO 336; 32pp + Sequence Listing; German.
XX
CC The present invention provides a number of human immune system associated
CC genes which are modified by the methylation of cytosines. The sequences
CC can be used in the diagnosis and treatment of immune system disorders,
CC including eye diseases such as retinopathy, neovascular glaucoma and
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC diseases. The present sequence is a gene of the invention.
XX
SQ Sequence 15674 BP; 4711 A; 271 C; 3032 G; 7660 T; 0 other;

Query Match 4.0%; Score 48.4; DB 24; Length 15674;
Best Local Similarity 48.2%; Pred. No. 0.23;
Matches 136; Conservative 0; Mismatches 146; Indels 0; Gaps 0;

Qy 185 TTGTGATTATGTCCTTCGTTGATTATTCTGTTTCATACAAGTCGTGTAATTTGCTGTTTG 244
Db 13049 TTTGGTTTTTATTTTTTTTTTTTTTATTTTATAATAGTAAAGTTGTTATAGTTTTTTTT 13108

Qy 245 TGACAGTACGATAGATCGACTCAACCTTCTGAGGTATTAGTTGAAGTTCATGTAAATTAG 304
Db 13109 TTATAGATAAATAAATTTAAAAAAATTATTTTTGTTATTTTTATAGATAAGGAAAGTGAG 13168

Qy 305 CTTTGTTTATCATAGTAGCATTTGATTATTGATGCTCTGTAGCTAATGATAAGCCATTGG 364
Db 13169 GTTTGGTTATAAGTTTAGGGAGGGACGGTGGTTGGATTGAAGTTTTTTTGATTTTATGGT 13228

Qy 365 AGGGAAGCAAGCTTTCTAAATGAATCTACGAATGGATGATAAAGTTCATGAATATTTTTG 424
Db 13229 TAGCGAAAAAATTTATAAATTTTTTTTTTAATGTATAAAAAATTTTTTTTGTTTTTATT 13288

Qy 425 TTACTTCTGCAGTCAGATCATGAGTTATTGAGTCTATTGTTT 466
Db 13289 TTTGTTTTAACGATATATTTATTTTTATTTTTTTTTTTATAT 13330

RESULT 12
ABL34477
ID ABL34477 standard; DNA; 15674 BP.
XX
AC ABL34477;
XX
DT 26-MAR-2002 (first entry)

```
XX
DE   Human metastasis associated gene SEQ ID NO: 30.
XX
KW   Metastasis associated gene; cytostatic; gene therapy; cancer;
KW   cytosine methylation; gene; ds.
XX
OS   Homo sapiens.
XX
PN   WO200177376-A2.
XX
PD   18-OCT-2001.
XX
PF   06-APR-2001; 2001WO-EP03970.
XX
PR   06-APR-2000; 2000DE-1019058.
PR   07-APR-2000; 2000DE-1019173.
PR   30-JUN-2000; 2000DE-1032529.
PR   01-SEP-2000; 2000DE-1043826.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A, Piepenbrock C, Berlin K;
XX
DR   WPI; 2002-010922/01.
XX
PT   New nucleic acid derived from chemically treated metastasis genes,
PT   useful for diagnosis of cancers by analysis of cytosine methylation,
PT   also for treatment -
XX
PS   Claim 1; SEQ ID NO 30; 23pp + Sequence Listing; English.
XX
CC   The present invention provides a number of human metastasis associated
CC   genes which are modified by cytosine methylation. The sequences can be
CC   used in the diagnosis and treatment of cancer. The present sequence is
CC   one of the genes of the invention.
XX
SQ   Sequence 15674 BP; 4711 A; 271 C; 3032 G; 7660 T; 0 other;

  Query Match             4.0%;  Score 48.4;  DB 24;  Length 15674;
  Best Local Similarity   48.2%;  Pred. No. 0.23;
  Matches  136;  Conservative    0;  Mismatches  146;  Indels    0;  Gaps    0;

Qy     185 TTGTGATTATGTCCTTCGTTGATTATTCTGTTTCATACAAGTCGTGTAATTTGCTGTTTG 244
Db   13049 TTTGGTTTTTATTTTTTTTTTTTTTTATTTTATAATAGTAAAGTTGTTATAGTTTTTTTT 13108

Qy     245 TGACAGTACGATAGATCGACTCAACCTTCTGAGGTATTAGTTGAAGTTCATGTAAATTAG 304
Db   13109 TTATAGATAAATAAATTTAAAAAAATTATTTTTGTTATTTTTATAGATAAGGAAAGTGAG 13168

Qy     305 CTTTGTTTATCATAGTAGCATTTGATTATTGATGCTCTGTAGCTAATGATAAGCCATTGG 364
Db   13169 GTTTGGTTATAAGTTTAGGGAGGGACGGTGGTTGGATTGAAGTTTTTTTGATTTTATGGT 13228

Qy     365 AGGGAAGCAAGCTTTCTAAATGAATCTACGAATGGATGATAAAGTTCATGAATATTTTTG 424
Db   13229 TAGCGAAAAAATTTATAAATTTTTTTTTTTAATGTATAAAAAATTTTTTTTGTTTTTATT 13288

Qy     425 TTACTTCTGCAGTCAGATCATGAGTTATTGAGTCTATTGTTT 466
Db   13289 TTTGTTTTAACGATATATTTATTTTTATTTTTTTTTTTATAT 13330


RESULT 13
AAL15210
ID   AAL15210 standard; cDNA; 883 BP.
XX
AC   AAL15210;
XX
DT   07-DEC-2001  (first entry)
XX
DE   Human breast cancer expressed polynucleotide 7667.
XX
KW   Human; breast cancer; cell marker; cytostatic; ss.
XX
OS   Homo sapiens.
XX
PN   WO200151628-A2.
XX
PD   19-JUL-2001.
XX
PF   10-JAN-2001; 2001WO-US00798.
XX
PR   14-JAN-2000; 2000US-0176077.
PR   14-MAR-2000; 2000US-0189167.
PR   24-MAR-2000; 2000US-0192099.
PR   29-MAR-2000; 2000US-0193480.
PR   15-MAY-2000; 2000US-0205230.
PR   09-JUN-2000; 2000US-0211315.
PR   25-JUL-2000; 2000US-0220534.
XX
PA   (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI   Lillie J, Xu Y, Wang Y, Steinmann K;
XX
DR   WPI; 2001-451856/48.
XX
PT   New peptide useful as a marker for the diagnosis of breast cancer -
XX
PS   Claim 1; Page 1378; 3695pp; English.
XX
CC   The invention relates to human breast cancer expressed polynucleotides
CC   (AAL07544-AAL26789) and methods of assessing whether a patient is
CC   afflicted with breast cancer by examining the correlation between the
CC   expression of certain markers and the cancerous state of breast cells.
CC   The polynucleotides and encoded polypeptides are potential markers for
CC   detecting, diagnosing, monitoring, characterising treating and
CC   potentially preventing breast cancer. The polynucleotides and encoded
CC   polypeptides are also useful for isolating compounds with cytostatic
CC   activity.
XX
SQ   Sequence 883 BP; 322 A; 18 C; 23 G; 396 T; 124 other;

  Query Match             3.8%;  Score 46.2;  DB 22;  Length 883;
  Best Local Similarity   39.1%;  Pred. No. 0.36;
  Matches  174;  Conservative    0;  Mismatches  271;  Indels    0;  Gaps    0;

Qy     223 AAGTCGTGTAATTTGCTGTTTGTGACAGTACGATAGATCGACTCAACCTTCTGAGGTATT 282
Db     312 AANTTATTTTATTTATTTAAANTAAAATTTTTNTNTTNANTTTTAAATTTNTNTTNNATA 371

Qy     283 AGTTGAAGTTCATGTAAATTAGCTTTGTTTATCATAGTAGCATTTGATTATTGATGCTCT 342
Db     372 ANTATAAAATTATTTTTTNAAAATTATTTTATTNNTNTTNTTTTAATTTATAGAANTAAA 431

Qy     343 GTAGCTAATGATAAGCCATTGGAGGGAAGCAAGCTTTCTAAATGAATCTACGAATGGATG 402
Db     432 AAAAATAAATTTTTTAATNTNNAANNNAAAAAAATTNTNTTTNAATTNAAGGTTTTAAAA 491

Qy     403 ATAAAGTTCATGAATATTTTTGTTACTTCTGCAGTCAGATCATGAGTTATTGAGTCTATT 462
Db     492 ANANTTTNAAAAAATTTTTTTTTTNTTTTNAGAAAAANANAAATAAAATTTTTTATAAANT 551

Qy     463 GTTTTTTTAAGCCTGTTTCAGATGATCCATCATCAGTAACAACATACACGGTGTAGTCCC 522
Db     552 ATAAAAAAAAAAAANNTTTTTTTTTAAATATNTTNATAAAAAAAAANTNNTTTTTTATTTAA 611

Qy     523 AAATCCATCATATGCACCTTCTTTTCTTCAATTTGGTCTTGTTTTTTTTTTTTTCATGATG 582
Db     612 AAANTTTTTTTATNNAAANAAAAAAAAAANTTGNTTAAAATTTNAAATTTTTTTTTTTTN 671

Qy     583 TCATTGAATTATTCAAGAAGTCACTTCGAGCATAATGATTTTTCAAAATCCACCTTTGTT 642
Db     672 ATTAAAAAAAATTNAANTTATTAANTTTTAAAAAATTTTTTTAAAAAAAAAATTTTGGTT 731

Qy     643 CAAGCACTACCACGTCTTTTCATCT 667
```

```
Db    732 TAATAANTTANAAAAANNTTTAATTT 756

RESULT 14
ABN80226
ID   ABN80226 standard; DNA; 19236 BP.
XX
AC   ABN80226;
XX
DT   15-JUL-2002  (first entry)
XX
DE   Human chemically modified disease associated gene SEQ ID NO 243.
XX
KW   Human; development; homeobox gene; HOX; diabetes; cancer; apoptosis;
KW   heart disease; epilepsy; histone deacetylation; muscular dystrophy;
KW   dwarfism; single nucleotide polymorphism; SNP; cytosine methylation;
KW   antidiabetic; cytostatic; anticonvulsant; ds.
XX
OS   Homo sapiens.
OS   Synthetic.
XX
PN   WO200200927-A2.
XX
PD   03-JAN-2002.
XX
PF   02-JUL-2001; 2001WO-EP07536.
XX
PR   30-JUN-2000; 2000DE-1032529.
PR   01-SEP-2000; 2000DE-1043826.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K;
XX
DR   WPI; 2002-130908/17.
XX
PT   Novel nucleic acid useful for diagnosis and therapy of diseases
PT   associated with development genes such as diabetes, comprises a
PT   sequence of a segment of chemically pretreated DNA of genes associated
PT   with development -
XX
PS   Claim 1; SEQ ID NO 243; 27pp; English.
XX
CC   The invention relates to a nucleic acid (I) comprising a sequence at
CC   least 18 bases in length of a segment of chemically pretreated DNA (II)
CC   of genes associated with development selected from 87 genes listed in
CC   the specification such as ACCPN, ADFN, or AFD1 and comprising one of 350
CC   sequences (ABN79984-ABN80333) or their complements. The invention is
CC   useful for the diagnosis or therapy of diseases associated with
CC   development genes, in particular disease related to homeobox containing
CC   genes (HOX), like diabetes, cancer, apoptosis related diseases, syndromes
CC   associated with congenital heart disease, epilepsy, diseases related to
CC   histone deacetylation, Currarino syndrome, diseases related with the
CC   development of the brain and limb girdle muscular dystrophy and dwarfism.
CC   Oligomers specific to each of the genes are useful for detecting the
CC   methylation state of all CpG dinucleotides within the 350 sequences or
CC   (II) and their complementary sequences, as primer oligonucleotides for
CC   the amplification of the 350 sequences, (II) and/or their complements and
CC   as oligomer probes for detecting the cytosine methylation state and/or
CC   single nucleotide polymorphisms (SNPs).
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification but is based on sequence information supplied to Derwent by
CC   the European Patent Office.
XX
SQ   Sequence 19236 BP; 5051 A; 361 C; 4619 G; 9205 T; 0 other;

  Query Match             3.8%;  Score 46.2;  DB 24;  Length 19236;
  Best Local Similarity   47.2%;  Pred. No. 0.79;
  Matches  141;  Conservative    0;  Mismatches  158;  Indels    0;  Gaps    0;

Qy    203 TTGATTATTCTGTTTCATACAAGTCGTGTAATTTGCTGTTTGTGACAGTACGATAGATCG 262
Db  17137 TTTATTCGTTTCGGTTTTTTAAGTTTTGGGATTATAGGTTTGAGTTATTATGTTTAGTCG 17196

Qy    263 ACTCAACCTTCTGAGGTATTAGTTGAAGTTCATGTAAATTAGCTTTGTTTATCATAGTAG 322
Db  17197 ATATAATTTTTTTTATTAATTGTTTGATGTTTTTTTAAAGTATTTTATTTTAGAGAGGTTT 17256

Qy    323 CATTTGATTATTGATGCTCTGTAGCTAATGATAAGCCATTGGAGGGAAGCAAGCTTTCTA 382
Db  17257 TTTTATTTTAAATATATTGTGTAATAAAAGTAAATATTTAGTAAAATTGATTTTTTTTG 17316

Qy    383 AATGAATCTACGAATGGATGATAAAGTTCATGAATATTTTTGTTACTTCTGCAGTCAGAT 442
Db  17317 TATAGTTTTAAGTTTAGATAAATATATAGATTTATGTAATTAGTTATTATATTGAATTTA 17376

Qy    443 CATGAGTTATTGAGTCTATTGTTTTTTTAAGCCTGTTTCAGATGATCCATCATCAGTAA 501
Db  17377 AGATAGAGAATCGTTTTTTAATTTTTATATATTTTTTTATGTTATTATTTTGTAGGTAA 17435

RESULT 15
ABL33748
ID   ABL33748 standard; DNA; 11805 BP.
XX
AC   ABL33748;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Human immune system associated gene SEQ ID NO: 1721.
XX
KW   Human; immune system disease; cytosine methylation; antiasthmatic;
KW   antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW   neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW   antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW   antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW   acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW   neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
KW   gene; ds.
XX
OS   Homo sapiens.
XX
PN   WO200200928-A2.
XX
PD   03-JAN-2002.
XX
PF   02-JUL-2001; 2001WO-EP07537.
XX
PR   30-JUN-2000; 2000DE-1032529.
PR   01-SEP-2000; 2000DE-1043826.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K;
XX
DR   WPI; 2002-130909/17.
XX
PT   Nucleic acid comprising fragment of chemically modified gene, useful
PT   for diagnosis and treatment of diseases associated with abnormal
PT   cytosine methylation -
XX
PS   Claim 1; SEQ ID NO 1721; 32pp + Sequence Listing; German.
XX
CC   The present invention provides a number of human immune system associated
CC   genes which are modified by the methylation of cytosines. The sequences
CC   can be used in the diagnosis and treatment of immune system disorders,
CC   including eye diseases such as retinopathy, neovascular glaucoma and
CC   macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC   leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC   rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC   diseases. The present sequence is a gene of the invention.
XX
SQ   Sequence 11805 BP; 3385 A; 75 C; 2168 G; 6177 T; 0 other;

  Query Match             3.8%;  Score 46;  DB 24;  Length 11805;
```

Best Local Similarity 43.2%; Pred. No. 0.78;
Matches 217; Conservative 0; Mismatches 285; Indels 0; Gaps 0;

```
Qy      85 TTCATCATCATACAGTATGGTTTTGATATGCTCTTCCATTATCACTGAGCCTTATGATTA 144
           || || ||  ||  || |   |||| |    | ||   ||   |  ||| |||| |
Db    9537 TTTATTATATTATTGTGTTTGTTTGTTTGTTTGTTTGTTTGAGATAGAGTTTTATTTTGT 9596

Qy     145 TGTTTTACGAGCTTATAATATCACTGATGGTGATTCAGTATTGTGATTATGTCCTTCGTT 204
           || ||     |       |        |||  || || | | |  |  || | |  || |||
Db    9597 TGGTTAGGTTGGAGTATAGTGGTACGATTTTGGTTTATTGTAATTTTTGTTTTTTTAGTT 9656

Qy     205 GATTATTCTGTTTCATACAAGTCGTGTAATTTGCTGTTTGTGACAGTACGATAGATCGAC 264
           |     | |||  |   |||  | | | | | || |  |   ||| |    ||
Db    9657 TAAGTGATTTTTTTGTTTTAGTTTTTTGAGTAGTTGGTATTATAGGTATGTGTTATTACG 9716

Qy     265 TCAACCTTCTGAGGTATTAGTTGAAGTTCATGTAAATTAGCTTTGTTTATCATAGTAGCA 324
           |  |  |  |    | |||  |  |  || |  |  |     | ||| ||| |
Db    9717 TTTAGTTAATTTTTTTTTATTATATTTTAAGTTTTAGGGTATATGTGTATAACGTGTAGG 9776

Qy     325 TTTGATTATTGATGCTCTGTAGCTAATGATAAGCCATTGGAGGGAAGCAAGCTTTCTAAA 384
           |||| |   |     | | |   | ||| |       ||| |   |   |  | || |  |
Db    9777 TTTGTTATATATGTATATATGTGTTATGTTGGTGTGTTGTATTTATTAATTCGTTATTTA 9836

Qy     385 TGAATCTACGAATGGATGATAAAGTTCATGAATATTTTTGTTACTTCTGCAGTCAGATCA 444
           |       ||   | |       |  |   | ||||| ||  || |  | | |  | |
Db    9837 ACGTTACGTATATTTTTTAATGTTATTTTTTTTTTTTTTTTTTATTTTATAATAGGTTTT 9896

Qy     445 TGAGTTATTGAGTCTATTGTTTTTTTAAGCCTGTTTCAGATGATCCATCATCAGTAACAA 504
           | ||       | | || ||| | |     |||||    || |  ||  | | | |  |
Db    9897 GGTGTGTGATGTTTTTTTTTTTGTGTTTATGTGTTTTTATTGTTTAATTTTTATTTATGA 9956

Qy     505 CATACACGGTGTAGTCCCAAATCCATCATATGCACCTTCTTTTCTTCAATTTGGTCTTGT 564
           | |   ||||||      |   |  | |   |  |  ||| || |   || |  | |
Db    9957 GTGAGAATATGTAGTGTTTCGTTTTTTGTTTTTGCGATAGTTTGTTGAGAATGATGGTTT 10016

Qy     565 TTTTTTTTTTTCATGATGTCAT 586
           ||  |||| || ||| | | ||
Db   10017 TTAGTTTTATTTATGTTTTTAT 10038
```

Search completed: May 24, 2003, 06:43:26
Job time : 437 secs

The Plant Cell, Vol. 10, 877-887, June 1998, www.plantcell.org © 1998 American Society of Plant Physiologists

RESEARCH ARTICLE

Transposon Tagging of the *Defective embryo and meristems* Gene of Tomato

James S. Keddie,[a,1,2] Bernard J. Carroll,[b,1] Colwyn M. Thomas,[c] Melquiades E. C. Reyes,[b] Victor Klimyuk,[c] Hans Holtan,[a] Wilhelm Gruissem,[a] and Jonathan D. G. Jones[c]

[a]Department of Plant Biology, 211 Koshland Hall, University of California, Berkeley, California 94720
[b]Departments of Biochemistry and Agriculture, University of Queensland, Brisbane 4072, Queensland, Australia
[c]Sainsbury Laboratory, John Innes Centre, Colney Lane, Norwich NR4 7UH, United Kingdom

The shoot and root apical meristems (SAMs and RAMs, respectively) of higher plants are mechanistically and structurally similar. This has led previously to the suggestion that the SAM and RAM represent modifications of a fundamentally homologous plan of organization. Despite recent interest in plant development, especially in the areas of meristem regulation, genes specifically required for the function of both the SAM and RAM have not yet been identified. Here, we report on a novel gene, *Defective embryo and meristems* (*Dem*), of tomato. This gene is required for the correct organization of shoot apical tissues of developing embryos, SAM development, and correct cell division patterns and meristem maintenance in roots. *Dem* was cloned using transposon tagging and shown to encode a novel protein of 72 kD with significant homology to YNV2, a protein of unknown function of *Saccharomyces cerevisiae*. *Dem* is expressed in root and shoot meristems and organ primordia but not in callus. The expression pattern of *Dem* mRNA in combination with the *dem* mutant phenotype suggests that *Dem* plays an important role within apical meristems.

INTRODUCTION

In plants, organogenesis is continuous and occurs in apices throughout the entire life cycle. This process is achieved by the action of apical meristems, which are groups of stem cells that are established early in embryogenesis and maintained in the tips of shoots and roots. Because apical meristems are almost entirely responsible for the elaboration of plant architecture, they have been a major subject of observational, experimental, and genetic studies (described in Steeves and Sussex, 1991; Meyerowitz, 1997). We are now beginning to elucidate the genes involved in meristem regulation and to understand their function (Meyerowitz, 1997).

In angiosperms, the shoot apical meristem (SAM) is usually a small dome of cells that consists of a peripheral zone in which leaves are initiated and a central zone in which the peripheral zone cells are replenished. The central zone contains cells that divide slowly, whereas the peripheral zone contains cells that divide rapidly (Lyndon, 1990; Steeves and Sussex, 1991). Superimposed upon this zonation are three clonally distinct cell layers (Poethig, 1987): L1 (forming the epidermis), L2 (forming the mesoderm), and L3 (forming the pith and vascular tissue). These cell layers generate the whole shoot. The L1 and L2 layers in the SAM are maintained by anticlinal cell divisions. Occasional cell divisions occur that result in the insertion of cells derived from one layer into the adjacent layer. These cells adopt a fate appropriate to their new layer, thus suggesting that positional information, rather than cell lineage, is the major factor influencing cell fate decisions during plant development. How cells in meristems communicate with each other has not yet been determined; however, recent results indicate roles for protein trafficking (Lucas et al., 1995) and extracellular signaling (Clark et al., 1997).

The root apical meristem (RAM), in contrast to the SAM, is an internal area of cells and is responsible for the production of cells for both the root and the root cap. The RAM is therefore surrounded on all sides by its derivatives. At the center of the root meristem is a region of cells known as the quiescent center—a population of cells that has a very long generation time. Surrounding the quiescent center are initial cells, which divide more rapidly and whose progeny differentiate into the basic cell types of the root and root cap. The cells of the quiescent center are proposed to act as replacements for the more rapidly dividing apical initials. Cell division patterns within the Arabidopsis root are almost invariant, which results in a root comprised of several clonally distinct

[1] These authors contributed equally to this work.
[2] Current address: Mendel Biotechnology Inc., 21375 Cabot Boulevard, Hayward, CA 94545. To whom correspondence should be addressed. E-mail jkeddie@mendelbio.com; fax 510-264-0254.

files of cells (Dolan et al., 1993). Lateral roots are not initiated at the root apex but rather are initiated from an internal layer of cells called the pericycle. Experimental evidence suggests that the root tip inhibits the formation of lateral roots (McCully, 1975).

Despite their differences, the basic organization of the SAM and RAM is similar: both meristems are layered structures that contain a central zone of quiescent or slowly dividing cells. In addition, experiments using surgically isolated meristems have shown that the SAM and RAM are autonomous in their development (Ball, 1952; Feldman and Torrey, 1976). These observations have led to the conclusion (Steeves and Sussex, 1991) that the differences between the SAM and RAM are superimposed upon a fundamentally homologous plan of organization and that the root and shoot systems probably represent evolutionary modifications of an "ancestral meristem" in response to different environments. Mutations that specifically affect both the SAM and RAM may therefore represent lesions in genes whose functions have been conserved throughout the evolution of apical meristems from the ancestral meristem.

In this study, we describe a recessive mutant of tomato, *defective embryo and meristems* (*dem*), that is affected in the development of both shoot and root apical meristems. *Dem* was cloned by using the transposable element *Dissociation* (*Ds*) as a tag and shown to encode a novel protein with a region of significant homology to a yeast protein of unknown function. *Dem* is expressed in SAMs and RAMs, axillary meristems, and organ primordia during adult plant growth. Although the exact function of *Dem* remains unclear, our initial observations suggest that it plays an important role within apical meristems and organ primordia.

RESULTS

dem Mutants Have Disrupted Apical Meristems

A total of 150 families carrying independent transpositions of the maize transposon *Ds* in tomato were generated. Approximately 25 seeds from each family were sown in flats, and seedlings were screened for mutant phenotypes. *dem* mutants were found in one family, N174 (Figure 1A). Testcross and F_2 analysis showed that the mutation was recessive and that mutant progeny occurred at a frequency of 10 to 15%. Self-pollination of heterozygotes revealed that the *dem* mutants had a highly variable number of small, slightly concave, abnormal cotyledons and no SAM. Wild-type seedlings were normally dicot (Figure 1B). Of 110 mutants inspected, two were monocot, 20 were dicot, 65 were tricot, and 23 were tetracot (Figures 1C to 1F).

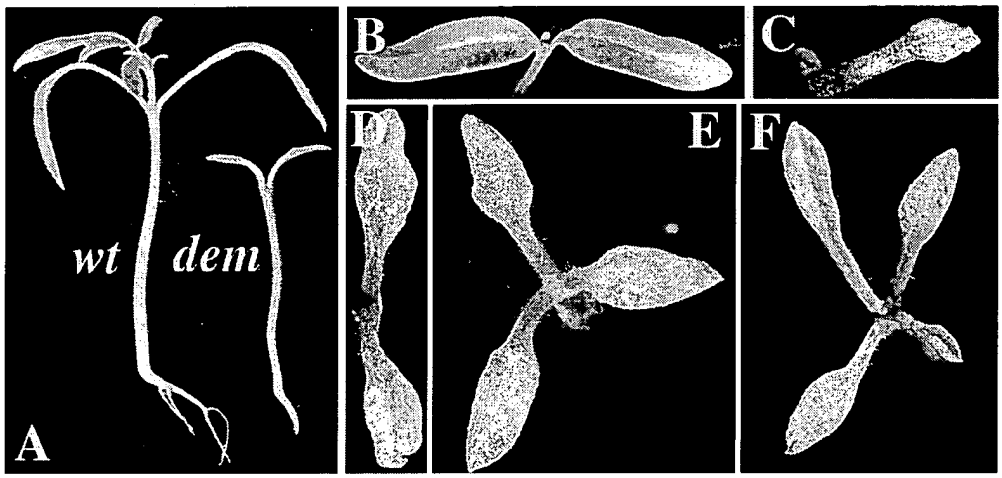

Figure 1. Seedling and Embryo Morphology Is Disrupted by the *dem* Mutation.

(A) Three-week-old *dem* and wild-type (*wt*) plants. *dem* plants have neither elongated roots nor a shoot.
(B) Wild-type dicot seedling.
(C) *dem* monocot.
(D) *dem* dicot.
(E) *dem* tricot.
(F) *dem* tetracot.

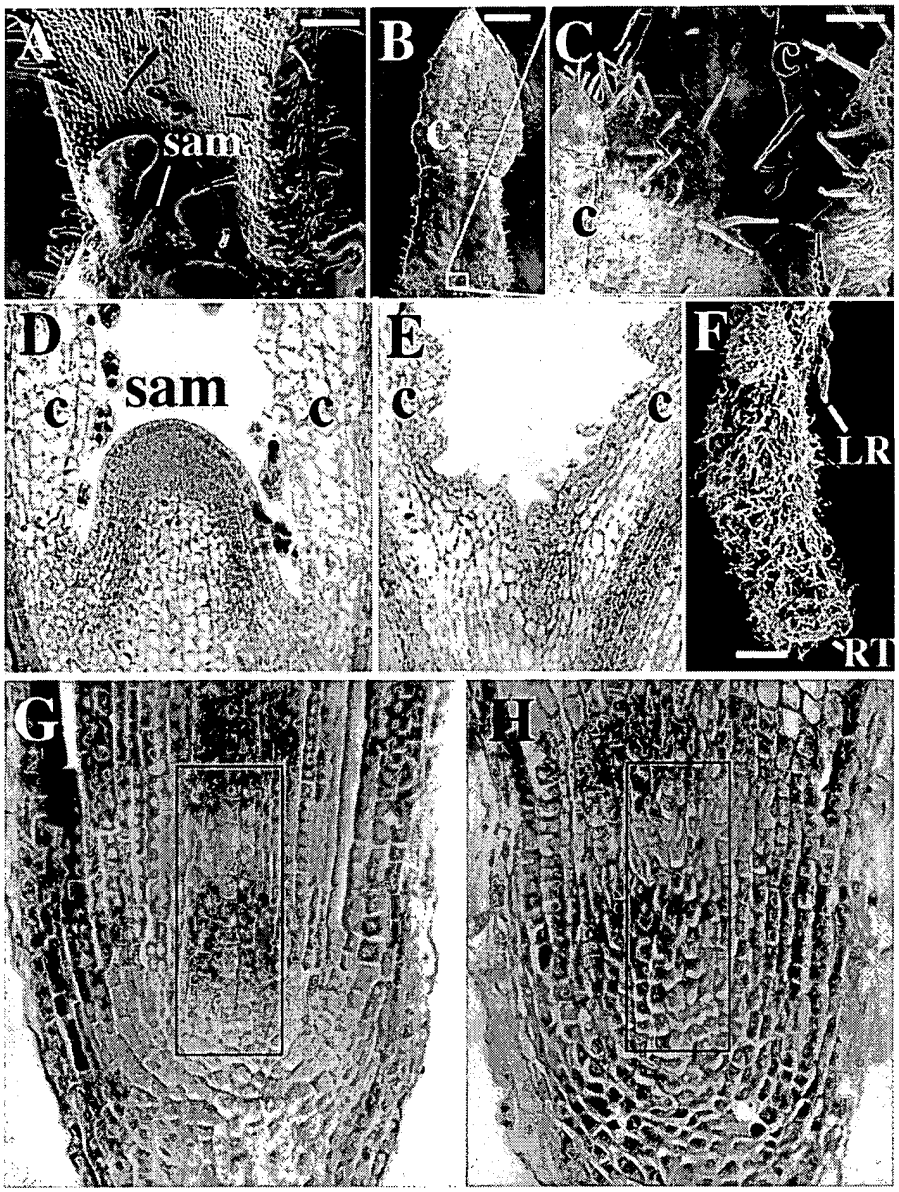

Figure 2. *dem* Seedlings Have No Apical Meristem.

(A) and **(B)** SEM of the SAM (sam) and a cotyledon (c) of a wild-type seedling and the cotyledon and shoot apical region of a *dem* mutant, respectively. One cotyledon has been cut off in **(A)** and **(B)** to facilitate viewing. Bar in **(A)** = 150 μm; bar in **(B)** = 719 μm.

(C) An expanded view of the *dem* apical region boxed in **(B)**. Bar = 76 μm.

(D) Section through the wild-type shoot apex.

(E) Section through the *dem* shoot apex. No typical SAM can be seen. The adaxial tissues of the cotyledons are disorganized.

(F) SEM of a *dem* root, showing a lateral root (LR) and root tip (RT). Bar = 712 μm.

(G) Section through the wild-type root apex, showing a typical root meristem.

(H) Section through the *dem* root apex, showing that the *dem* root meristem (boxed) is disorganized.

Apical growth of *dem* seedlings was terminated soon after germination, and no true leaves were initiated (Figure 2). Scanning electron microscopy (SEM) studies (Figures 2A to 2C) showed that the apical region between the cotyledons of dicot *dem* seedlings usually contained no SAM or leaf primordia. Sections through *dem* apices (Figures 2D and 2E) confirmed that no organized SAM was present in *dem* seedlings but rather that tissue with a disorganized cell arrangement formed. This disorganization continued from the axis of the cotyledons into the adaxial half of the cotyledons. Cell organization in the abaxial half of cotyledons appeared to be normal. *dem* roots terminated after 3 or 4 mm of growth, and lateral roots, which also aborted after a short period of extension (Figure 2F), were initiated. *dem* roots were also very hairy; however, it is not possible to predict whether this is a direct effect of the mutation. Sections through a *dem* root show that although many of the outer cell files are correctly maintained, cells in the center of the root apex are disorganized compared with the wild type (boxed in Figures 2G and 2H). No clear cell files were observed in the central cylinder of a *dem* root.

Isolation of the *Dem* Gene by Transposon Tagging

Sequences flanking the *Ds* element in a *dem* mutant were cloned using inverse polymerase chain reaction (IPCR) (Thomas et al., 1994) and sequenced. Using this sequence, two primers, dem3′ and dem5′, were designed. When used in combination with primer B34 (Thomas et al., 1994), they could be used to map the *Ds* element in relation to the *dem* phenotype. In tests of 200 individuals of a segregating population using triplex PCR, *Ds* was found to segregate with the *dem* phenotype (Figure 3A), demonstrating close linkage between the mutant phenotype and a *Ds* insertion. A transposase source, stabilized *Activator* (*sAc*), was crossed onto a *dem* heterozygote, and an F_1 plant containing both *sAc* and *Ds* was self-pollinated. Approximately 75% of the *dem*

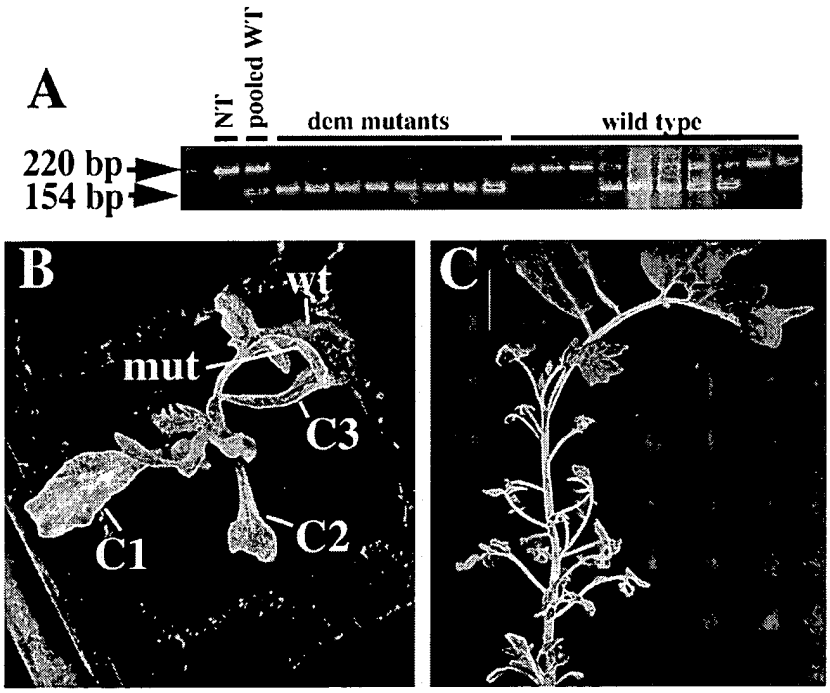

Figure 3. Linkage of *Ds* to the *dem* Mutation and Somatic Reversion of *dem*.

(A) Linkage of the *Ds* insertion to the *dem* mutation was demonstrated using a PCR zygosity test: a 220-bp fragment was amplified from the preinsertion allele, and a 154-bp fragment was amplified from the *Ds* insertion allele. PCR with DNA from stable mutant seedlings only produced a 154-bp fragment, indicating that these seedlings are homozygous for the *Ds* insertion and that the *Ds* is closely linked to the *dem* mutation. Wild-type plants were either heterozygous or homozygous for the preinsertion allele. NT, untransformed; WT, wild type.

(B) and **(C)** Transposase-dependent somatic reversion of the *dem* phenotype confirmed that the *Ds* insertion is the cause of the *dem* mutation. For further details, see Methods and Results. Somatic revertants also initiated leaves that could not develop properly. mut, mutant; wt, wild type; C1, C2, and C3 indicate cotyledons.

mutants in this segregating population contained *sAc*. After a period of up to 2 months, all of these mutants reverted and formed shoots from between their cotyledons. These shoots were either fully wild type or chimeric. The chimeric shoots contained both wild-type and mutant tissues and were rather unusual in appearance (Figures 3B and 3C). After a period of time, chimeric shoots became fully wild type in appearance. In contrast, *dem* mutants that did not contain *sAc* never formed shoots, even after several months. The shoots of somatic revertants yielded fruit that contained viable seed. Seeds from somatic revertants were planted, and seedlings were scored for the *dem* phenotype: the majority of these seedlings were wild type, demonstrating that the *dem* mutation is germinally unstable in the presence of the transposase gene. The close linkage of *Ds* with the *dem* phenotype and the *sAc*-dependent somatic and germinal instability of the *dem* phenotype strongly implicate *Ds* as the cause of the *dem* mutation.

An 8-bp target site duplication is typical of *Ds* insertion, and many *Ds* excision alleles retain this duplication or have deletions/substitutions of one or two nucleotides (Saedler and Nevers, 1985). To confirm that the *dem* mutation was caused by a *Ds* insertion, DNA from germinal revertants was prepared, and the sequence alterations expected from *Ds* excision were analyzed. All sequenced *Dem* revertant alleles contained sequence alterations consistent with *Ds* excision (Figure 4A). This result confirms that the *dem* mutation is a result of a *Ds* insertion into the *Dem* locus. The *Ds* insertion allele of *dem* was designated dem^{Ds}.

During the course of the analysis of germinal revertants, a sAc^{-} Ds^{-} plant was identified that gave rise to ~10% mutant progeny. This allele of *dem* was later sequenced and found to contain a 7-bp insertion at the *Ds* insertion site that causes an early frameshift in the *Dem* open reading frame (ORF). This allele was designated dem^{+7}. Plants homozygous for dem^{+7} displayed a phenotype identical to dem^{Ds}, demonstrating that dem^{Ds} is probably a null allele. The phenotypic analysis described above was performed with mutants homozygous for dem^{Ds}.

In a separate experiment, dem^{+7} heterozygotes containing *sAc* were crossed onto dem^{Ds} heterozygotes, and several somatic revertants were identified. DNA was extracted from wild-type tissues of these mutants. The sequences surrounding the site of *Ds* insertion were amplified by polymerase chain reaction (PCR) using oligonucleotides dem3′ and dem5′, with one being kinase labeled. PCR products were then size fractionated by PAGE. All revertant alleles represented either perfect excision events or insertions/deletions of +3, +6, or −3 nucleotides (Figure 4B). These sequence alterations restored the *Dem* reading frame and resulted in the addition or loss of one or two amino acids in the Dem protein. In one case, a deletion of 18 nucleotides (leading to a deletion of six amino acids in the Dem protein) was identified. These results are consistent with the idea that *Ds* insertion occurred in the *Dem* coding sequence and that only excision events that do not alter the reading frame will reinstate wild-type gene function. Amino acid residues around this area are therefore not essential for the function of the Dem protein.

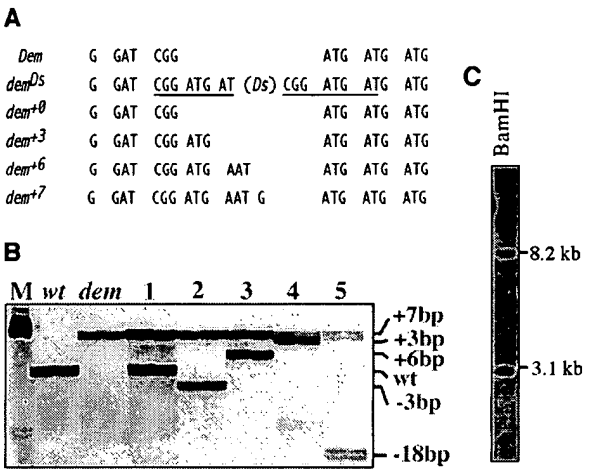

Figure 4. *dem* Excision Alleles.

(A) *Ds* insertion into *Dem* creates an 8-bp direct repeat (underlined). *dem* excision alleles containing in-frame insertions (+6 and +3) and wild-type sequence all restored wild-type gene function. dem^{+7} is a stable allele of *dem* containing a 7-bp insertion. dem^{+7} is predicted to produce a peptide of 123 amino acids before translation is terminated.

(B) Gel analysis of excision alleles, showing that footprints of −3, +3, +6, and −18 (lanes 2, 3, 4, and 5, respectively) reinstate wild-type (wt) gene activity. These revertants contain the dem^{+7} mutant allele and the revertant wild-type allele. M indicates length markers.

(C) Blot of BamHI-digested tomato genomic DNA hybridized with the *Dem* cDNA. A BamHI restriction site exists within the *Dem* cDNA; therefore, two bands of 8.2 and 3.1 kb indicate one gene.

PCR tests showed that dem^{Ds} is fully transmitted through male and female gametes. The observed segregation distortion (10 to 15% mutant rather than 25% mutant) is due to decreased viability of *dem* embryos (M.E.C. Reyes and B.J. Carroll, unpublished data). DNA gel blotting experiments using low- and high-stringency washes demonstrated that *Dem* is present as a singly copy in the tomato genome (Figure 4C).

Dem Encodes a Novel Protein

Cloning and sequencing of the flanking DNA of this mutant line revealed that the *Ds* element had inserted in a large ORF. The cloned flanking sequences were used to screen a cDNA library. One full-length *Dem* cDNA clone was isolated and sequenced (Figure 5; GenBank accession number

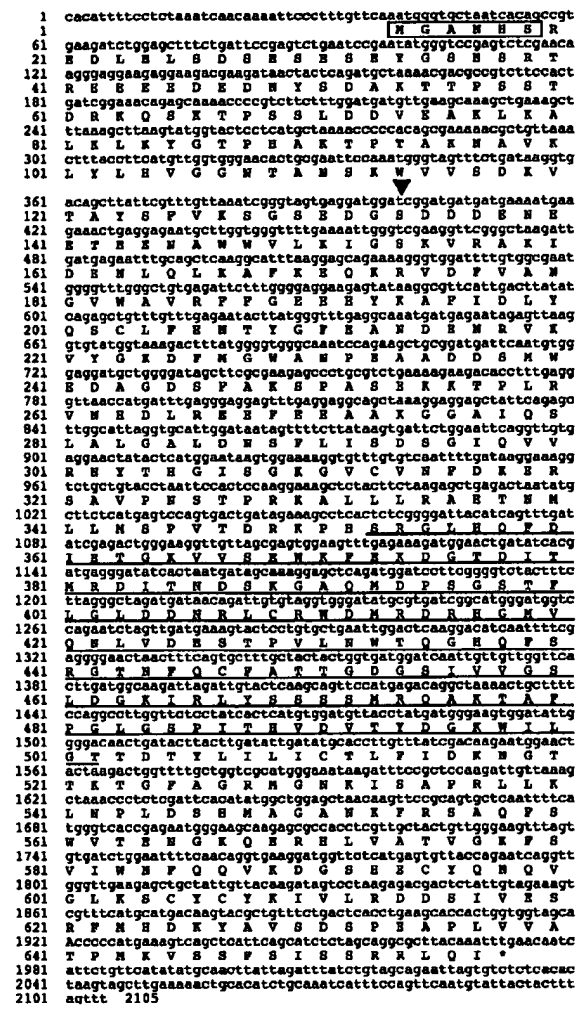

Figure 5. *Dem* cDNA and Deduced Amino Acid Sequence.

The DNA sequence of the longest *Dem* cDNA is shown, with the predicted amino acid sequence provided at the bottom. The region of homology to YNV2 is underlined, and a potential myristoylation motif is boxed. *Ds* was inserted into codon 121 of *Dem* in Dem^{Ds} plants (indicated by a filled triangle). The GenBank accession number is Y13632.

Y13632). This cDNA contained one long ORF with an in-frame stop codon in the 5′ leader sequence. Translation of the ORF predicted a charged protein of 71,919 D with a pI of 5.58 (Figure 5).

A search of the PROSITE database showed that the predicted mature N-terminal sequence of Dem, MGANHS, conforms to the consensus sequence for N-myristoylation, suggesting that Dem may be attached by a lipid anchor to a cellular membrane. BLAST (Altschul et al., 1997) searches using the Dem peptide sequence identified two potentially homologous proteins (Figure 6A): CYPRO4 from artichoke thistle (GenBank accession number P40781; 93% identical and 98% similar; $P = 5.4 \times 10^{-260}$) and YNV2 from *Saccharomyces cerevisiae* (GenBank accession number P40157; 33% identical and 51% similar over 150 residues; $P = 1.7 \times 10^{-11}$). Both proteins are of unknown function. Two Arabidopsis expressed sequence tags (F19919 and N96644) with strong homology to the 3′ and 5′ ends of *Dem* were also identified (Figure 6B).

Dem Is Expressed in Apical Meristems and Organ Primordia

The expression pattern of *Dem* was investigated initially using RNA gel blot hybridization analysis (Figure 7). *Dem* mRNA is expressed at high levels in shoot apices and at much lower levels in roots, young fruit, stem, mature leaves, and seedlings. Importantly, no *Dem* transcript was identified in RNA prepared from callus, suggesting that Dem is not a component of the cell cycle machinery and is not required for cell maintenance or unpatterned cell division.

Dem expression was further localized by in situ RNA hybridization (Figures 8A to 8M). In shoot apices, *Dem* expression was restricted to apical meristems and adaxial sides of leaf primordia (Figures 8A, 8C, and 8D) and young leaves (Figure 8F), which corresponds closely to the tissues affected in *dem* seedlings. *Dem* was downregulated in mature leaf tissue and upregulated in the adaxial side of the leaf in the region of developing leaflet primordia (Figure 8G). *Dem* was also expressed in dormant axillary meristems (Figure 8I) and in floral meristems and developing flowers (Figure 8J). In root tips, *Dem* mRNA formed a gradient that was most concentrated at the root apex (Figure 8L). Overall, *Dem* was expressed in tissues of adult plants in which organized cell division occured and in vascular strands. No signal was observed in sense strand controls (Figures 8B, 8E, 8H, 8K, and 8M).

DISCUSSION

Dem is expressed in all regions of the plant in which organized cell divisions take place. These regions include apical meristems, organ primordia, and leaflet primordia. However, *Dem* is not expressed in callus. Furthermore, loss of *Dem* function causes disorganization of both the shoot and root apex and in the adaxial tissues of cotyledons. These observations suggest that *Dem* is required for the organization or maintenance of meristems and primordia. *dem* mutants are morphologically distinct from those previously reported to be affected in basic body planning (Jurgens et al., 1991; Mayer et al., 1991), SAM development (Caruso, 1968; Meyerowitz, 1997), and root development (Benfey and Schiefelbein, 1994) and may represent a novel category of mutants that are affected in a basic aspect of meristem regulation.

A

```
Dem 340  SRGLHQFDIETGKVVSEWKFEKDGTDITMRDITNDSKGAQMDPSGSTFLGLDDNRLCRWDMRDRHGMVQNLVDESTPV
         .. |..|||| |||.||.    ..| .: :.. ..| .||.|.. |::|:.:. : : | |  :|. .  ||||.
YNV2 491 ENKLYKMDIERGKVIEEWS....TGDKNVVQYGPTKKFDQMTPEQ.TIVGVSQKGVFKIDPRI.NGKNKIAVDESK..

Dem 440  LNWTQGHQFSRGTNFQCFATTGDGSIVVGSLDGKIRLYSSSSMRQAKTAFPGLGSPITHVDVTYDGKWILGT 489
                 ::    || :::||:.| |.:|| .|.|:||.. ::| ||||:|:||| :|. :... ||||:|:|
YNV2 576 .......DYVGKYNFSSIGTTESGYIAIGSEKGDIKLYDRLGIR.ARTAIPSLGQAIKFITTSADGKWLLAT 624
```

B

```
Dem 563      VGKFSVIWNFQQVKDGSHECYQNQVGLKSCYCYKIVL..RDDSIVESRFMHDKYAVS 599
             VGKFSVIW+ ++VK+ +HECY+NQ GLKSCYC KI++  +D+SIVESRFMHD ++ S
F19919       VGKFSVIWDLERVKNSAHECYRNQQGLKSCYCNKILVLLKDESIVESRFMHDNFSFS

Dem  56      SSLDDVEAKLKALKLKYGTPHAKTPTAKNAVKLYLHVGGNTANSKWVVSDKVTA 109
             SSLDDVEAKL+ALKLKY    +  P+ +N+ +L+ ++ GNT  +KWV ++K+TA
N96644       SSLDDVEAKLQALKLKYPLTQSA.PSTQNSARLFRYINGNTPKAKWVTAEKLTA
```

Figure 6. Sequence Alignments.

(A) Sequence alignment of Dem with YNV2 from yeast. Vertical lines show identical amino acids, and single and double dots represent similar amino acids, according to the Genetics Computer Group (Madison, WI) BESTFIT program.

(B) Amino acid sequence alignments of *Dem* with two Arabidopsis expressed sequence tags. F19919 is the 3′ end sequence, whereas N96644 is the 5′ end sequence.

The shoot apex of *dem* embryos lacks a shoot meristem and has a highly variable number of cotyledons that contain disorganized cells. During normal embryogenesis, two cotyledons are initiated on the apical flanks of a globular stage embryo, and the SAM becomes morphologically apparent between the emerging cotyledons (Jurgens et al., 1991; Mayer et al., 1991). In this context, we foresee a role for *Dem* in the organized cell divisions that accompany the transition between the globular and heart stages of embryogenesis. Without correct cell divisions, the normal signals that coordinate cotyledon number and apical development may be perturbed. This would lead to a deregulation of cotyledon number, altered cell division planes, and the failure to form or maintain a SAM. Because the true relationship between the cotyledons and the SAM in wild-type plants is not known (Kaplan, 1969; Barton and Poethig, 1993; Endrizzi et al., 1996), it is impossible to predict whether the SAM simply is not established in *dem* embryos or is established but not maintained.

In the *dem* root apex, cell divisions within a central zone of cells are disorganized, and several of the central cell files in the root are not correctly formed. *dem* roots terminate after a short period of growth but have the ability to initiate determinate lateral roots. The simplest explanation of the *dem* root phenotype is that the *dem* mutation makes roots determinate; primary roots and lateral roots can be formed, but they cannot be maintained. The fact that both terminal and lateral root meristems are determinate in *dem* mutants suggests that the *dem* SAM may also be determinate. These observations suggest that *Dem* may play a role in meristem maintenance and that the *dem* SAM is consumed in embryogenesis during the formation of cotyledons.

A possible clue for *Dem* function comes from the observation that two other mutants with altered cotyledon number have altered hormone levels. The Arabidopsis mutant *pinoid* (Bennett et al., 1995) frequently produces tricot seedlings and has defects in auxin transport. The mutant *altered meristem program1* also has a highly variable cotyledon number and altered cytokinin levels. Furthermore, polar auxin transport has been shown to be critical for pattern formation during embryogenesis (Liu et al., 1993), leaf and floral organ phyllotaxy (Meicenheimer, 1981; Okada et al., 1991), compound leaf development (Avasarala et al., 1996), meristem maintenance (Avasarala et al., 1996), and root meristem organization (Kerk and Feldman, 1995). Thus, *Dem* may be involved in cellular responses to hormone gradients that organize all apices and organ primordia.

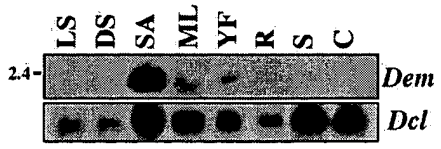

Figure 7. Tissue Distribution of *Dem* mRNA.

RNA gel blot analysis of *Dem* expression in light-grown seedlings (LS), dark-grown seedlings (DS), shoot apices (SA), mature leaves (ML), young fruit (YF), roots (R), stem (S), and callus (C). RNA was hybridized using the entire *Dem* cDNA, washed, and exposed for 48 hr. This blot was previously hybridized with a probe to *Dcl*, a constitutively and ubiquitously expressed mRNA (Keddie et al., 1996). A single message of 2.4 kb was observed.

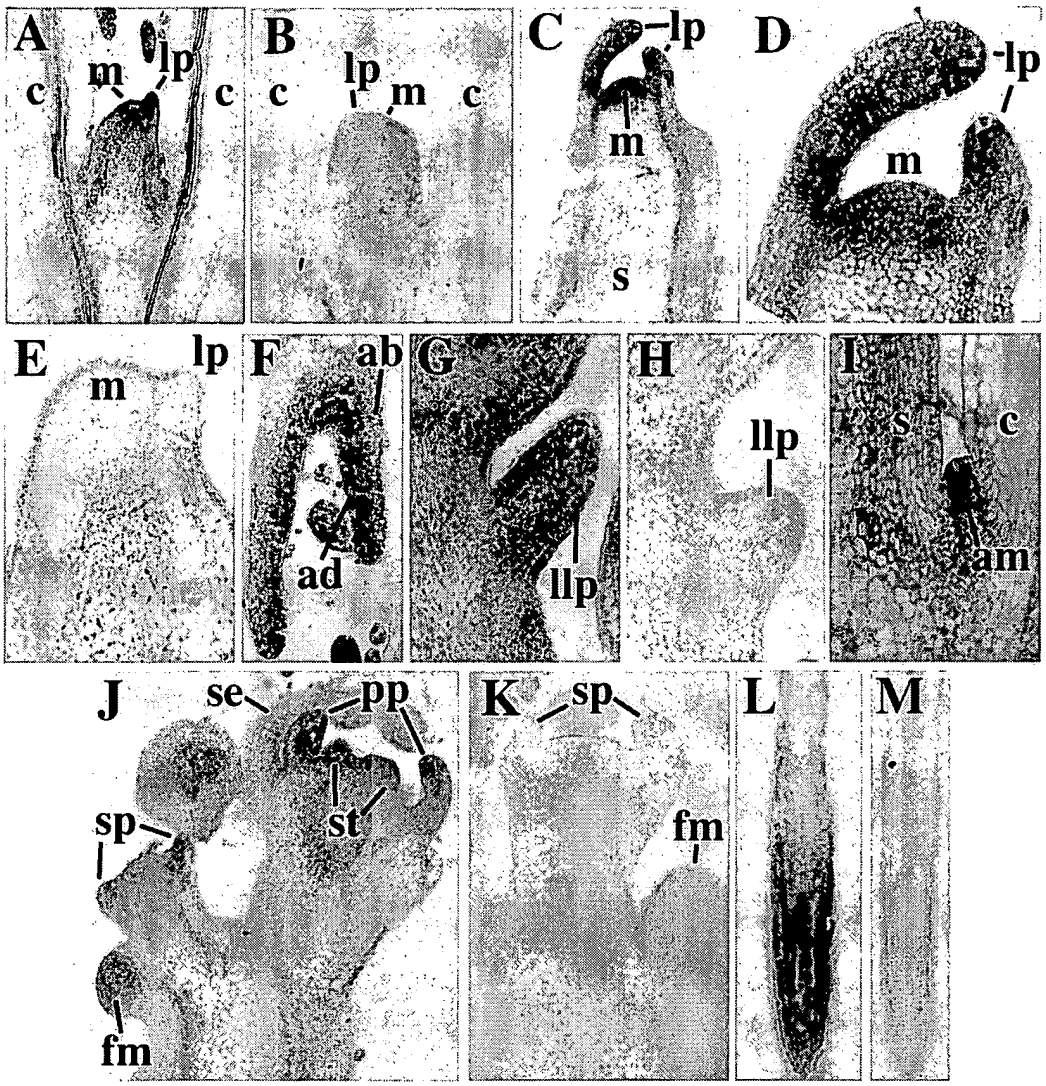

Figure 8. In Situ Distribution of *Dem* mRNA.

(A) and **(B)** SAM and cotyledons of 12-day-old seedlings. Signal was observed in the meristem and leaf primordia, using antisense **(A)** but not sense **(B)** *Dem* probes.

(C) to **(E)** SAM and stem from 4-week-old plants. Signal was observed in meristems and leaf primordia, using antisense **(C)** and **(D)** but not sense **(E)** probes.

(F) Cross-section showing *Dem* expression in the adaxial tissues of young leaves.

(G) and **(H)** *Dem* expression was detected in leaflet primordia, using sense **(G)** but not antisense **(H)** probes.

(I) Stem/cotyledon axis. *Dem* expression is detected in axillary meristems.

(J) and **(K)** Cross-section through inflorescence showing an emerging floral meristem and developing flowers. Staining is observed in floral meristems and in organ primordia as they emerge, using *Dem* antisense **(J)** but not *Dem* sense **(K)** probes.

(L) and **(M)** Sections through root tips. Staining was observed in root tips, using antisense **(L)** but not sense **(M)** probes.

Digoxigenin labeling is visible as brown staining. All hybridizations used 10-μm-thick sections and digoxigenin-labeled *Dem* probe. ab, abaxial; ad, adaxial; am, axillary meristem; c, cotyledon; fm, floral meristem; llp, leaflet primordium; lp, leaf primordia; m, meristem; pp, petal primordium; s, stem; se, sepal; sp, sepal primordium; st, stamen primordium.

A notable feature of the expression pattern of *Dem*, at least in shoot apices, is that it is apparently coincident with the expression of *tKn1*, a gene encoding a KNOTTED1-related homeodomain protein of tomato (Hareven et al., 1996). KNOTTED1-related proteins are believed to maintain cells in an undifferentiated state within meristems (Smith et al., 1992) and in the leaf and leaflet primordia of tomato (Hareven et al., 1996). Similar to *Dem*, *knotted1*-related genes of maize are not expressed in callus tissue (Smith et al., 1992) and are expressed in vascular strands (Smith et al., 1992; Jackson et al., 1994). Also, in Arabidopsis, mutations in *STM*, a *Knotted1* homolog, result in seedlings with no apparent SAM (Long et al., 1996). These observations suggest that *Dem* may be required for correct cell division patterns within the domain of *Knotted* expression.

In summary, we have identified a mutant, *dem*, that plays an important role in the maintenance or function of both the SAM and RAM. We have cloned the *Dem* gene by transposon tagging and shown that it is expressed in all areas of the plant in which organized cell division is taking place. The conceptual translation of the *Dem* cDNA provides little evidence regarding the function of the Dem protein. The lack of apparent nuclear localization sequences or DNA binding motifs suggests that it is not a nuclear transcription factor. The presence of myristoylation consensus motifs makes it tempting to speculate that Dem may be anchored to a cellular membrane. The homology of Dem to a yeast protein raises the possibility that Dem is a cellular component that has evolved to become an essential gene for organized cell divisions that occur in meristems and primordia during plant development.

METHODS

Transgenic Plant Material and Generation of the *defective embryo and meristems* Mutant

Transgenic tomato (*Lycopersicon esculentum*) cultivar Moneymaker carrying maize transposable elements was used for all experiments. A total of 150 transposants was generated from a single *Dissociation* (*Ds*) T-DNA line (1561E) by selection for excision and reinsertion of *Ds* after testcrossing a 1561E/10512I double heterozygote to wild-type plants (Carroll et al., 1995). The 10512I line carries the transposase gene (*sAc*) linked to β-glucuronidase (GUS). Seedlings carrying a transposed *Ds* were self-pollinated, and the progeny were screened for mutations. Family N174 carries a single transposed *Ds* and includes mutants exhibiting the *defective embryo and meristems* (*dem*) phenotype.

Reversion of the *dem* Mutant in the Presence of the Transposase

To demonstrate instability of *dem* in the presence of a transposase, a *Dem* heterozygote was crossed to the transposase line 10512I (Carroll et al., 1995). F_1 double heterozygotes for the *Ds* insertion and the transposase gene were identified by a polymerase chain reaction (PCR) test (identifying dem^{Ds}; see below) and histochemical staining for GUS (the marker for the transposase gene). F_1 double heterozygotes were selfed, and the F_2 generation was screened for GUS-positive mutant seedlings. GUS-positive mutants were observed for somatic instability of the mutant phenotype. Somatic F_2 revertants were testcrossed to an untransformed tester, and the progeny were screened for germinal wild-type and mutant excision alleles at the *Dem* locus, as described below.

Cloning the *Dem* cDNA

Fragments of the *Dem* gene were cloned by inverse PCR (IPCR) (Thomas et al., 1994) and used to screen a λgt10 cDNA library constructed using seedling mRNA. We purified six positives from 5×10^5 plaques, and one full-length *Dem* cDNA was sequenced on both strands. RNA and genomic DNA extraction and analysis were performed as described previously (Keddie et al., 1996).

PCR Test for *Ds* Zygosity at the *Dem* Locus

The mutant line was maintained as a heterozygote. To detect zygosity for the *Ds* insertion in *Dem*, we developed a simple triplex PCR test (Thomas et al., 1994) with intact leaf tissue (Klimyuk et al., 1993; Carroll et al., 1995). Based on the sequences flanking both sides of the *Ds* in *dem*, oligonucleotide primers dem5′ (5′-TTTCTGCTCCTTAAATGCATTGAG-3′) and dem3′ (5′-TTCATGTTGGTGGGAACACTGCGA-3′) were designed to amplify a 220-bp preinsertion fragment. dem5′, in combination with primer B34 (5′-ACGGTCGGTACGGGATTTTCCCAT-3′), which primes from sequences at the end of *Ds*, amplifies a 154-bp fragment corresponding to the *Ds* insertion in the *dem* gene. By using PCR with these three primers, we performed zygosity tests for the *Ds* insertion in *Dem* on individual seedlings.

PCR Footprint Analysis of *Dem* Revertants

Footprint analysis was done using oligonucleotides dem5′ and dem3′. PCR products from wild-type and germinal revertant plants were cloned and sequenced. In addition, a screen for new excision alleles was performed by crossing sAc^+ $dem^{+/?}$ heterozygotes with dem^{Ds} heterozygotes. The seeds from this cross were germinated, and wild-type plants were discarded. After ~1 month, ~50% of the mutants initiated a shoot from between their cotyledons, and growth was resumed. To analyze the size of footprints left in new *dem* alleles, either dem3′ or dem5′ was kinase labeled with γ-labeled ^{33}P-ATP and used with the other primer to amplify excision alleles. PCR products were denatured and separated on a 6% polyacrylamide gel (Figure 4B).

Microscopy and in Situ Hybridization

Samples for light microscopy were prepared using a microwave procedure. Tissue was fixed twice in formaldehyde acetic acid (FAA) at 37°C for a total of 30 min, dehydrated at 67°C in 70% ethanol and then 100% ethanol for 75 sec each, treated in 2-propanol at 75°C for 90 sec, and then embedded in molten Paraplast (Pelco, Reading, CA) at 67°C for ~3 hr in a 3440 MAX Laboratory microwave (Pelco). A full version of this protocol can be obtained from the National Science

Foundation Center home page (www.plantbio.berkeley.edu). Samples were serially sectioned, stained in safranin O and orange gold to highlight densely cytoplasmic cells, and viewed on an Axiophot microscope (Carl Zeiss, Inc., Thornwood, NY). Samples prepared for scanning electron microscopy (SEM) were fixed in FAA, dehydrated in ethanol, dried in a critical point dryer, sputter coated with palladium to 20 nm, and viewed on a DS130 scanning electron microscope (ISI, Philadelphia, PA).

In situ RNA hybridization was performed using methods described by Coen et al. (1990). An internal 559-bp EcoRI fragment of the *Dem* cDNA was subcloned into pBluescript SK+ (Stratagene, La Jolla, CA). T7- and T3-primed digoxigenin-labeled RNA probes were made using digoxigenin RNA labeling mix (Boehringer Mannheim) and hydrolyzed at 60°C for 30 min in 100 mM carbonate buffer, pH 10.2. A minimum of three samples were examined per experiment, and sense strand controls were always included.

ACKNOWLEDGMENTS

We thank the staff of the University of California at Berkeley National Science Foundation Center for Plant Developmental Biology for training, colleagues at the University of California at Berkeley for stimulating discussion about meristems, and members of the Gruissem laboratory for critical reading of the manuscript and advice. J.S.K. was supported by a Human Frontiers Science Program Long-Term Fellowship. B.J.C. is grateful to the University of Queensland for an Australian Research Council grant. Research in the laboratory of W.G. is supported by the National Science Foundation, and research in the Sainsbury Laboratory is funded by the Gatsby Charitable Trust.

Received January 13, 1998; accepted March 20, 1998.

REFERENCES

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. Nucleic Acids Res. **25,** 3389–3402.

Avasarala, S., Yang, J., and Caruso, J.L. (1996). Production of phenocopies of the *Lanceolate* mutant in tomato using auxin transport inhibitors. J. Exp. Bot. **47,** 709–712.

Ball, E. (1952). Morphogenesis of shoots after isolation of the shoot apex of *Lupinus albus*. Am. J. Bot. **39,** 167–191.

Barton, M.K., and Poethig, R.S. (1993). Formation of the shoot apical meristem in *Arabidopsis thaliana*: An analysis of development in the wild-type and in the *SHOOT MERISTEMLESS* mutant. Development **119,** 823–831.

Benfey, P.N., and Schiefelbein, J.W. (1994). Getting to the root of plant development: The genetics of *Arabidopsis* root formation. Trends Genet. **10,** 84–88.

Bennett, S.R.M., Alvarez, J., Bossinger, G., and Smyth, D.R. (1995). Morphogenesis in *pinoid* mutants of *Arabidopsis*. Plant J. **8,** 841–857.

Carroll, B.J., Klimyuk, V.I., Thomas, C.M., Bishop, G.J., Harrison, K., Scofield, S., and Jones, J.D.G. (1995). Germinal transpositions of the maize element *Dissociation* from T-DNA loci in tomato. Genetics **139,** 407–420.

Caruso, J.L. (1968). Morphogenetic aspects of a leafless mutant in tomato. I. General patterns of development. Am. J. Bot. **55,** 1169–1176.

Clark, S.E., Williams, R.W., and Meyerowitz, E.M. (1997). The *CLAVATA1* gene encodes a putative receptor kinase that controls shoot and floral meristem size in Arabidopsis. Cell **89,** 575–585.

Coen, E.S., Romero, J.M., Doyle, S., Elliot, R., Murphy, G., and Carpenter, R. (1990). Floricaula: A homeotic gene required for flower development in *Antirrhinum majus*. Cell **63,** 1311–1322.

Dolan, L., Janmaat, K., Willemsen, V., Linstead, P., Poethig, S., Roberts, K., and Scheres, B. (1993). Cellular organization of the *Arabidopsis thaliana* root. Development **119,** 71–84.

Endrizzi, K., Moussian, B., Haecker, A., Levin, J.Z., and Laux, T. (1996). The *SHOOT MERISTEMLESS* gene is required for maintenance of undifferentiated cells in *Arabidopsis* shoot and floral meristems and acts at a different regulatory level than the meristem genes *WUSCHEL* and *ZWILLE*. Plant J. **10,** 967–979.

Feldman, L.J., and Torrey, J.G. (1976). The isolation and culture *in vitro* of the quiescent center of *Zea mays*. Am. J. Bot. **52,** 580–590.

Hareven, D., Gutfinger, T., Parnis, A., Eshed, Y., and Lifschitz, E. (1996). The making of a compound leaf: Genetic manipulation of leaf architecture. Cell **84,** 735–744.

Jackson, D., Veit, B., and Hake, S. (1994). Expression of maize *KNOTTED1* related homeobox genes in the shoot apical meristem predicts patterns of morphogenesis in the vegetative shoot. Development **120,** 405–413.

Jurgens, G., Mayer, U., Torres-Ruiz, R.A., Berleth, T., and Misera, S. (1991). Genetic analysis of pattern formation in the *Arabidopsis* embryo. Development **113** (suppl.), 27–38.

Kaplan, D. (1969). Seed development in *Downingia*. Phytomorphology **19,** 153–178.

Keddie, J.S., Carroll, B., Jones, J.D.G., and Gruissem, W. (1996). The *DCL* gene of tomato is required for chloroplast development and palisade cell morphogenesis in leaves. EMBO J. **15,** 4208–4217.

Kerk, N.M., and Feldman, J.L. (1995). A biochemical model for the initiation and maintenance of the quiescent center: Implications for organization of root meristems. Development **121,** 2825–2883.

Klimyuk, V., Carroll, B.J., Thomas, C.M., and Jones, J.D.G. (1993). Alkali treatment for rapid preparation of plant tissue for reliable PCR analysis. Plant J. **3,** 493–494.

Liu, C.-m., Xu, Z.-h., and Chua, N.-H. (1993). Auxin polar transport is essential for the establishment of bilateral symmetry during early plant embryogenesis. Plant Cell **5,** 621–630.

Long, J.A., Moan, E.I., Medford, J.I., and Barton, M.K. (1996). A member of the KNOTTED class of homeodomain proteins encoded by the *STM* gene of *Arabidopsis*. Nature **353,** 402–407.

Lucas, W.J., Bouche-Pillon, S., Jackson, D., Nguyen, L., Baker, L., Ding, B., and Hake, S. (1995). Selective trafficking of KNOTTED1 homeodomain protein and its mRNA through plasmodesmata. Science **270,** 1980–1983.

Lyndon, R.F. (1990). Plant development: The cellular basis. In Topics in Plant Physiology, M. Black and J. Chapman, eds (London: Unwin Hyman), pp. 181–238.

Mayer, U., Torres-Ruiz, R.A., Berleth, T., Misera, S., and Jurgens, G. (1991). Mutations affecting body organization in the *Arabidopsis* embryo. Nature **353,** 402–407.

McCully, M.E. (1975). The development of lateral roots. In The Development and Function of Roots, J.G. Torrey and D.T. Clarkson, eds (New York: Academic Press), pp. 39–58.

Meicenheimer, R.D. (1981). Changes in *Epilobium* phyllotaxy induced by *N*-1-naphthylphthalamic acid and α-4-chlorophenoxy-isobutyric acid. Am. J. Bot. **68,** 1139–1154.

Meyerowitz, E.M. (1997). Genetic control of cell division patterns in developing plants. Cell **88,** 299–308.

Okada, K., Ueda, J., Komaki, M.K., Bell, C.J., and Shimura, Y. (1991). Requirement of the auxin polar transport system in early stages of *Arabidopsis* floral bud formation. Plant Cell **3,** 677–684.

Poethig, R.S. (1987). Clonal analysis of cell lineage patterns in plant development. Am. J. Bot. **74,** 581–594.

Saedler, H., and Nevers, P. (1985). Transposition in plants: A molecular model. EMBO J. **4,** 585–590.

Smith, L., Greene, B., Veit, B., and Hake, S. (1992). A dominant mutation in the maize homeobox gene, *Knotted-1,* causes its ectopic expression in leaf cells with altered fates. Development **116,** 21–30.

Steeves, T.A., and Sussex, I.M. (1991). Patterns in Plant Development. (Cambridge, UK: Cambridge University Press).

Thomas, C.M., Jones, D.A., English, J., Carroll, B.J., Bennetzen, J.L., Harrison, K., Burbridge, A., Bishop, G.J., and Jones, J.D.G. (1994). Analysis of the chromosomal distribution of transposase-carrying T-DNAs in tomato using the inverse polymerase chain reaction. Mol. Gen. Genet. **242,** 573–585.